Journal of Medical Internet Research

Impact Factor (2018): 4.945 - ranked #1 medical informatics journal by Impact Factor
Volume 21 (2019), Issue 1 ISSN: 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Electronic Systems for Patients to Report and Manage Side Effects of Cancer Treatment: Systematic Review (e10875)
Lorraine Warrington, Kate Absolom, Mark Conner, Ian Kellar, Beverly Clayton, Michael Ayres, Galina Velikova. . . . 6

Consumer Health Information Technology in the Prevention of Substance Abuse: Scoping Review (e11297)
Apoorva Pradhan, Leah Park, Fadia Shaya, Joseph Finkelstein. . . . 30

Web-Based Digital Health Interventions for Weight Loss and Lifestyle Habit Changes in Overweight and Obese Adults: Systematic Review and Meta-Analysis (e298)
Alline Beleigoli, Andre Andrade, Alexandre Cançado, Matheus Paulo, Maria Diniz, Antonio Ribeiro. . . . 118

Identifying Brief Message Content for Interventions Delivered via Mobile Devices to Improve Medication Adherence in People With Type 2 Diabetes Mellitus: A Rapid Systematic Review (e10421)
Hannah Long, Yvonne Bartlett, Andrew Farmer, David French. . . . 129

Internet-Delivered Acceptance and Commitment Therapy for Anxiety Treatment: Systematic Review (e12530)
Joshua Kelson, Audrey Rollin, Brad Ridout, Andrew Campbell. . . . 172

Evaluating Digital Maturity and Patient Acceptability of Real-Time Patient Experience Feedback Systems: Systematic Review (e9076)
Mustafa Khanbhai, Kelsey Flott, Ara Darzi, Erik Mayer. . . . 237

Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration (e12959)
Bhone Kyaw, Nakul Saxena, Pawel Posadzki, Jitka Vseteckova, Charoula Nikolaou, Pradeep George, Ushashree Divakar, Italo Masiello, Andrzej Kononowicz, Nabil Zary, Lorainne Tudor Car. . . . 623

Viewpoints

E-Learning for Medical Education in Sub-Saharan Africa and Low-Resource Settings: Viewpoint (e12449)
Sandra Barteit, Albrecht Jahn, Sekelani Banda, Till Bärnighausen, Annel Bowa, Geoffrey Chileshe, Dorota Guzek, Margarida Jorge, Sigrid Lüders, Gregory Malunga, Florian Neuhann. . . . 51

Use of the Principles of Design Thinking to Address Limitations of Digital Mental Health Interventions for Youth: Viewpoint (e11528)
Hanneke Scholten, Isabela Granic. . . . 57

Measuring the Implementation of Behavioral Intervention Technologies: Recharacterization of Established Outcomes (e11752)
Eric Hermes, Aaron Lyon, Stephen Schueller, Joseph Glass. . . . 71

Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line (e11507)
Anthony Pisani, Nitya Kanuri, Bob Filbin, Carlos Gallo, Madelyn Gould, Lisa Lehmann, Robert Levine, John Marcotte, Brian Pascal, David Rousseau, Shairi Turner, Shirley Yen, Megan Ranney. . . . 465

Tutorial

Human-Centered Design of Video-Based Health Education: An Iterative, Collaborative, Community-Based Approach (e12128)
Maya Adam, Shannon McMahon, Charles Prober, Till Bärnighausen. . . . 85

Original Papers

Web-Based Exercise as an Effective Complementary Treatment for Patients With Nonalcoholic Fatty Liver Disease: Intervention Study (e11250)
Daniel Pfirrmann, Yvonne Huber, Jörn Schattenberg, Perikles Simon. . . . 100

Evaluating a Web-Based Social Anxiety Intervention Among Community Users: Analysis of Real-World Data (e11566)
Hugh McCall, Fjola Helgadottir, Ross Menzies, Heather Hadjistavropoulos, Frances Chen. . . . 148

A Novel Insight Into the Challenges of Diagnosing Degenerative Cervical Myelopathy Using Web-Based Symptom Checkers (e10868)
Benjamin Davies, Colin Munro, Mark Kotter. . . . 161

Electronic Health Program to Empower Patients in Returning to Normal Activities After Colorectal Surgical Procedures: Mixed-Methods Process Evaluation Alongside a Randomized Controlled Trial (e10674)
Chantal den Bakker, Judith Huirne, Frederieke Schaafsma, Charlotte de Geus, Hendrik Bonjer, Johannes Anema. . . . 188

Using the Facebook Advertisement Platform to Recruit Chinese, Korean, and Latinx Cancer Survivors for Psychosocial Research: Web-Based Survey Study (e11571)
William Tsai, Daisy Zavala, Sol Gomez. . . . 201

How Search Engine Data Enhance the Understanding of Determinants of Suicide in India and Inform Prevention: Observational Study (e10179)
Natalia Adler, Ciro Cattuto, Kyriaki Kalimeri, Daniela Paolotti, Michele Tizzoni, Stefaan Verhulst, Elad Yom-Tov, Andrew Young. . . . 211

Association Between Cancer Incidence and Mortality in Web-Based Data in China: Infodemiology Study (e10677)
Chenjie Xu, Yi Wang, Hongxi Yang, Jie Hou, Li Sun, Xinyu Zhang, Xinxi Cao, Yabing Hou, Lan Wang, Qiliang Cai, Yaogang Wang. . . . 221

Additional Telemedicine Rounds as a Successful Performance-Improvement Strategy for Sepsis Management: Observational Multicenter Study (e11161)
Robert Deisz, Susanne Rademacher, Katrin Gilger, Rudolf Jegen, Barbara Sauerzapfe, Christina Fitzner, Christian Stoppe, Carina Benstoem, Gernot Marx. . . . 247

Use of the CHA2DS2-VASc Score for Risk Stratification of Hospital Admissions Among Patients With Cardiovascular Diseases Receiving a Fourth-Generation Synchronous Telehealth Program: Retrospective Cohort Study (e12790)
Jen-Kuang Lee, Chi-Sheng Hung, Ching-Chang Huang, Ying-Hsien Chen, Pao-Yu Chuang, Jiun-Yu Yu, Yi-Lwun Ho. 256

A Web-Based Application to Improve Data Collection in an Interventional Study Targeting Childhood Obesity: Pre-Post Analysis (e10861)
Meagan Hanbury, Banafsheh Sadeghi, Iraklis Tseregounis, Rosa Gomez-Camacho, Rosa Manzo, Maria Rangel, Bogdan Alexandrescu, Adela de la Torre. 267

Digital Exclusion Among Mental Health Service Users: Qualitative Investigation (e11696)
Ben Greer, Dan Robotham, Sara Simblett, Hannah Curtis, Helena Griffiths, Til Wykes. 275

Differences Between Mothers and Fathers of Young Children in Their Use of the Internet to Support Healthy Family Lifestyle Behaviors: Cross-Sectional Study (e11454)
Rachel Laws, Adam Walsh, Kylie Hesketh, Katherine Downing, Konsita Kuswara, Karen Campbell. 284

How Women Use Digital Technologies for Health: Qualitative Interview and Focus Group Study (e11481)
Deborah Lupton, Sarah Maslen. 296

Internet Access and Hypertension Management Among the Elderly Population: A Nationally Representative Cross-Sectional Survey in China (e11280)
Yinzi Jin, Mingxia Jing, Luyu Zhang, Suhang Song, Xiaochen Ma. 307

Technical Support by Smart Glasses During a Mass Casualty Incident: A Randomized Controlled Simulation Trial on Technically Assisted Triage and Telemedical App Use in Disaster Medicine (e11939)
Andreas Follmann, Marian Ohligs, Nadine Hochhausen, Stefan Beckers, Rolf Rossaint, Michael Czaplik. 318

Informing Adults With Back Pain About Placebo Effects: Randomized Controlled Evaluation of a New Website With Potential to Improve Informed Consent in Clinical Research (e9955)
Felicity Bishop, Maddy Greville-Harris, Jennifer Bostock, Amy Din, Cynthia Graham, George Lewith, Christina Liossi, Tim O'Riordan, Peter White, Lucy Yardley. 328

Improving Electronic Health Record Note Comprehension With NoteAid: Randomized Trial of Electronic Health Record Note Comprehension Interventions With Crowdsourced Workers (e10793)
John Lalor, Beverly Woolf, Hong Yu. 344

EuroQol (EQ-5D-5L) Validity in Assessing the Quality of Life in Adults With Asthma: Cross-Sectional Study (e10178)
Gimena Hernandez, Olatz Garin, Alexandra Dima, Angels Pont, Marc Martí Pastor, Jordi Alonso, Eric Van Ganse, Laurent Laforest, Marijn de Bruin, Karina Mayoral, Vicky Serra-Sutton, Montse Ferrer, ASTRO-LAB Group. 357

A Modular Health-Related Quality of Life Instrument for Electronic Assessment and Treatment Monitoring: Web-Based Development and Psychometric Validation of Core Thrive Items (e12075)
Paul Wicks, Stacey McCaffrey, Kim Goodwin, Ryan Black, Michael Hoole, James Heywood. 369

Improving the Efficacy of Cognitive Training for Digital Mental Health Interventions Through Avatar Customization: Crowdsourced Quasi-Experimental Study (e10133)
Max Birk, Regan Mandryk. 388

Assessing the Effectiveness of Engaging Patients and Their Families in the Three-Step Fall Prevention Process Across Modalities of an Evidence-Based Fall Prevention Toolkit: An Implementation Science Study (e10008)
Megan Duckworth, Jason Adelman, Katherine Belategui, Zinnia Feliciano, Emily Jackson, Srijesa Khasnabish, I-Fong Lehman, Mary Lindros, Heather Mortimer, Kasey Ryan, Maureen Scanlan, Linda Berger Spivack, Shao Yu, David Bates, Patricia Dykes. 406

Farm Owners and Workers as Key Informants in User-Centered Occupational Health Prototype Development: A Stakeholder-Engaged Project (e9711)
Bryan Weichelt, Casper Bendixsen, Matthew Keifer. . . . 417

The Absence of Evidence is Evidence of Non-Sense: Cross-Sectional Study on the Quality of Psoriasis-Related Videos on YouTube and Their Reception by Health Seekers (e11935)
Simon Mueller, Pierre Jungo, Lucian Cajacob, Simon Schwegler, Peter Itin, Oliver Brandt. . . . 428

Consumer Health Search on the Web: Study of Web Page Understandability and Its Integration in Ranking Algorithms (e10986)
Joao Palotti, Guido Zuccon, Allan Hanbury. . . . 437

Leading by Example: Web-Based Sexual Health Influencers Among Men Who Have Sex With Men Have Higher HIV and Syphilis Testing Rates in China (e10171)
Dan Wu, Weiming Tang, Haidong Lu, Tiange Zhang, Bolin Cao, Jason Ong, Amy Lee, Chuncheng Liu, Wenting Huang, Rong Fu, Katherine Li, Stephen Pan, Ye Zhang, Hongyun Fu, Chongyi Wei, Joseph Tucker. . . . 475

Digital Recruitment and Acceptance of a Stepwise Model to Prevent Chronic Disease in the Danish Primary Care Sector: Cross-Sectional Study (e11658)
Lars Larsen, Jens Sondergaard, Janus Thomsen, Anders Halling, Anders Sønderlund, Jeanette Christensen, Trine Thilsing. . . . 493

Expertise Modulates Students' Perception of Pain From a Self-Perspective: Quasi-Experimental Study (e10885)
Sareh Said Yekta-Michael, André Schüppen, Arnim Gaebler, Jens Ellrich, Jan Koten. . . . 510

Identifying Common Methods Used by Drug Interaction Experts for Finding Evidence About Potential Drug-Drug Interactions: Web-Based Survey (e11182)
Amy Grizzle, John Horn, Carol Collins, Jodi Schneider, Daniel Malone, Britney Stottlemyer, Richard Boyce. . . . 520

Measuring the Impact of an Open Web-Based Prescribing Data Analysis Service on Clinical Practice: Cohort Study on NHS England Data (e10929)
Alex Walker, Helen Curtis, Richard Croker, Seb Bacon, Ben Goldacre. . . . 534

Application of Efficient Data Cleaning Using Text Clustering for Semistructured Medical Reports to Large-Scale Stool Examination Reports: Methodology Study (e10013)
Hyunki Woo, Kyunga Kim, KyeongMin Cha, Jin-Young Lee, Hansong Mun, Soo Cho, Ji Chung, Jeung Pyo, Kun-Chul Lee, Mira Kang. . . . 543

Use of In-Game Rewards to Motivate Daily Self-Report Compliance: Randomized Controlled Trial (e11683)
Sara Taylor, Craig Ferguson, Fengjiao Peng, Magdalena Schoeneich, Rosalind Picard. . . . 549

Measurement of Quality of Life in Patients with Mycosis Fungoides/Sézary Syndrome Cutaneous T-Cell Lymphoma: Development of an Electronic Instrument (e11302)
Stacey McCaffrey, Ryan Black, Mitchell Nagao, Marjan Sepassi, Gaurav Sharma, Susan Thornton, Youn Kim, Julia Braverman. . . . 565

An Electronic Patient-Reported Outcome Tool for the FACT-B (Functional Assessment of Cancer Therapy-Breast) Questionnaire for Measuring the Health-Related Quality of Life in Patients With Breast Cancer: Reliability Study (e10004)
Lina Matthies, Florin-Andrei Taran, Lucia Keilmann, Andreas Schneeweiss, Elisabeth Simoes, Andreas Hartkopf, Alexander Sokolov, Christina Walter, Nina Sickenberger, Stephanie Wallwiener, Manuel Feisst, Paul Gass, Michael Lux, Florian Schuetz, Peter Fasching, Christof Sohn, Sara Brucker, Joachim Graf, Markus Wallwiener. . . . 578

The Most Influential Medical Journals According to Wikipedia: Quantitative Analysis (e11429)
Dariusz Jemielniak, Gwinyai Masukume, Maciej Wilamowski. . . . 594

Perceptions of Visualizing Physical Activity as a 3D-Printed Object: Formative Study (e12064)
Sam Crossley, Melitta McNarry, Joanne Hudson, Parisa Eslambolchilar, Zoe Knowles, Kelly Mackintosh. 605

Server-Focused Security Assessment of Mobile Health Apps for Popular Mobile Platforms (e9818)
Jannis Müthing, Raphael Brüngel, Christoph Friedrich. 636

Short Paper

Assessing the Impact of a Social Marketing Campaign on Program Outcomes for Users of an Internet-Based Testing Service for Sexually Transmitted and Blood-Borne Infections: Observational Study (e11291)
Mark Gilbert, Travis Salway, Devon Haag, Michael Kwag, Joshua Edward, Mark Bondyra, Joseph Cox, Trevor Hart, Daniel Grace, Troy Grennan, Gina Ogilvie, Jean Shoveller. 486

Review

Electronic Systems for Patients to Report and Manage Side Effects of Cancer Treatment: Systematic Review

Lorraine Warrington[1], BSc, MSc, PhD; Kate Absolom[1], BSc, PhD; Mark Conner[2], BSc, PhD; Ian Kellar[2], BA, D Phil; Beverly Clayton[1], RGN, RSCN, BHSc; Michael Ayres[3], MBBS, MSc; Galina Velikova[1], BMBS, PhD

[1]Section of Patient Centred Outcomes Research, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, United Kingdom

[2]School of Psychology, University of Leeds, Leeds, United Kingdom

[3]Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom

Corresponding Author:
Lorraine Warrington, BSc, MSc, PhD
Section of Patient Centred Outcomes Research
Leeds Institute of Cancer and Pathology
University of Leeds
Beckett Street
Leeds, LS9 7TF
United Kingdom
Phone: 44 1132068504
Email: l.warrington@leeds.ac.uk

Abstract

Background: There has been a dramatic increase in the development of electronic systems to support cancer patients to report and manage side effects of treatment from home. Systems vary in the features they offer to patients, which may affect how patients engage with them and how they improve patient-centered outcomes.

Objective: This review aimed to (1) describe the features and functions of existing electronic symptom reporting systems (eg, symptom monitoring, tailored self-management advice), and (2) explore which features may be associated with patient engagement and patient-centered outcomes.

Methods: The review was registered with the International Prospective Register of Systematic Reviews (PROSPERO) and followed guidelines from the Centre for Reviews and Dissemination (University of York, United Kingdom). Primary searches were undertaken of MEDLINE, Embase, PsycInfo, Web of Science, Cochrane Central Register of Controlled Trials, and the Health Technology Assessment databases. Secondary searches were undertaken by screening reference lists and citations. Two researchers applied broad inclusion criteria to identify and select relevant records. Data were extracted and summarized using Microsoft Excel. In order to meet the aims, the study selection, data extraction, and data synthesis comprised two stages: (1) identifying and characterizing available systems and (2) summarizing data on patient engagement and patient-centered outcomes.

Results: We identified 77 publications relating to 41 distinct systems. In Stage 1, all publications were included (N=77). The features identified that supported clinicians and care were facility for health professionals to remotely access and monitor patient-reported data (24/41, 58%) and function to send alerts to health professionals for severe symptoms (17/41, 41%). Features that supported patients were facility for patients to monitor/review their symptom reports over time (eg, graphs) (19/41, 46%), general patient information about cancer treatment and side effects (17/41, 41%), tailored automated patient advice on symptom management (12/41, 29%), feature for patients to communicate with the health care team (6/41, 15%), and a forum for patients to communicate with one another (4/41, 10%). In Stage 2, only publications that included some data on patient engagement or patient-centered outcomes were included (N=29). A lack of consistency between studies in how engagement was defined, measured, or reported, and a wide range of methods chosen to evaluate systems meant that it was not possible to compare across studies or make conclusions on relationships with system features.

Conclusions: Electronic systems have the potential to help patients manage side effects of cancer treatment, with some evidence to suggest a positive effect on patient-centered outcomes. However, comparison across studies is difficult due to the wide range of assessment tools used. There is a need to develop guidelines for assessing and reporting engagement with systems, and a set of core outcomes for evaluation. We hope that this review will contribute to the field by introducing a taxonomy for characterizing system features.

Trial Registration: PROSPERO CRD42016035915; www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016035915

(J Med Internet Res 2019;21(1):e10875) doi:10.2196/10875

KEYWORDS

oncology; chemotherapy; patient reported outcomes; patient centered; medical informatics

Introduction

Increased efficacy of cancer treatments has led to a rising global population of people living with and beyond cancer. Effective multimodal cancer treatments can slow disease progression, ease the symptoms of the disease, and in some cases cure disease altogether. However, treatments can cause a vast array of side effects such as nausea, pain, fatigue, and diarrhea, which may negatively affect a patient's quality of life (QoL) and may even become life-threatening, with severe cases such as neutropenic infections. Many cancer treatments are delivered in an ambulatory setting and methods of follow-up and support are highly variable dependent on disease, treatment type, and local practice and resources. Information is commonly provided by the health care team on expected and possible side effects, and patients are advised to seek help if symptoms become a cause for concern. However, patients may not always be able to fully absorb this information at the time it is provided [1] or feel confident in making decisions on when additional hospital contact is necessary between routine clinical reviews [2]. Furthermore, clinicians are mainly reliant on interpreting patient retrospective reports of treatment side effects to ensure safety of care and manage supportive medications. Side effects are not often documented in medical records in a consistent and comparable way [3].

Over the past decade, there has been a dramatic increase in the number of electronic systems developed to support patients during and after cancer treatment by using patient-reported outcome measures (PROMs) to remotely assess symptoms [4-8]. The routine use of PROMs in oncology care as a strategy to enhance symptom monitoring has demonstrated many benefits, such as improved communication between clinicians and patients, and better symptom awareness [9]. Using electronic systems to collect and manage PROMs data has the potential to overcome some of the common challenges previously associated with collating data collected on paper. More recently developed systems can be accessed from any Web-enabled device, allowing patients to report symptoms from home using their own electronic devices such as computers, tablets, or mobile phones. This can be done in real time, rather than relying on retrospective reporting and potentially allows automated documentation of patient reports in the medical record [10].

There is considerable variation in the features offered by symptom reporting systems. Some primarily focus on making symptom data routinely available to health professionals and provide alerts when severe symptoms have been reported [5,11-15]. Others have been developed with a greater focus on patient self-management, delivering tailored and automated self-management advice when appropriate, and advising patients to contact their health care team when necessary [8,16-20]. Some systems use a combination of both approaches [4] and can also include additional features such as facilitating communication with medical teams or other patients.

The availability or absence of certain features may affect how patients engage with systems [21,22]. The terms "engagement" and "adherence" are often used interchangeably in this context. However, adherence suggests an optimal way to use a technology and this is not always easy to define [23]. For the purposes of this review, we refer to engagement in a broad sense of levels of patient usage of the technology. Understanding the key components that can enhance patient engagement with electronic symptom reporting is potentially crucial for improving the development of future systems and encouraging their implementation into standard practice. There are many factors that are likely to have an impact, from individual differences [24], socioeconomic status and healthy literacy [25], to basic system usability [21,26]. There is relatively little currently known about the underlying processes and particularly the role that the availability of systems features might play. However, there is evidence to suggest that individuals vary in the features that they value and use most [20]. In addition, needs may change over time, as patients become more experienced with the system, but also with their disease and treatment [27].

The presence or absence of system features is also likely to affect the level of patient benefit gained from using the system. For example, changes in behavior or disease outcome have been more often observed with interactive interventions in comparison with those that are purely educational [28]. While the use of interactive online systems is associated with greater self-efficacy, better self-management, and more participation in health care [29-32], this may be associated only with specific features such as interactive communication and progress tracking features [33], and consultation and self-management support [34].

Systematic reviews traditionally focus on high-quality evidence for a specific research question. However, increasingly, the value of taking a broader approach to inclusion is being recognized as important to answer complex research questions, particularly in the emerging field of online health interventions [35,36]. With this in mind, the focus of this review was to take an inclusive approach to systematically review and describe the features and functions of existing systems. We also wanted to focus on understanding the level of evidence indicating whether key system features are associated with better patient system engagement and patient-centered outcomes.

The aims of this systematic review are to (1) describe the features and functions of existing electronic symptom reporting systems developed for patients during cancer treatment, and (2) explore which features of these systems may be associated with patient engagement and outcomes. Specifically, we wanted to summarize (1) patient engagement and whether this is related

to specific system features (eg, symptom monitoring, tailored self-management advice), and (2) patient-centered outcomes used to evaluate systems and whether better outcomes are associated with specific features.

Methods

Protocol and Registration

Details of the protocol were registered on the International Prospective Register of Systematic Reviews (PROSPERO) database [37]. There were no major deviations from the protocol. However, study selection, data extraction, and data synthesis comprised two stages: (1) identifying and characterizing available systems, and (2) summarizing data on patient engagement and patient-centered outcomes. This staged approach was not initially planned but was necessary in order to meet the aims of the review.

Eligibility Criteria

The review question was refined using Population, Intervention, Comparator, Outcomes, Study design (PICOS) criteria (Table 1), and eligibility criteria were developed based on this. For Stage 1, we wanted an overview of all systems available, so all relevant publications including published abstracts, protocols, and qualitative studies were included. However, discussion papers or systematic reviews were excluded. For Stage 2, in order to review evidence available on patient engagement and any patient-centered outcomes, we wanted to include feasibility studies with any evaluation data of patient use, rather than restricting criteria to randomized controlled trials (RCTs) only. Only full papers were included in this stage. Criteria were piloted by 2 researchers (LW and KA) on a subset of 10 randomly selected papers and subsequently refined and clarified before the next stage.

Information Sources

Studies were identified from systematic searches of Medline, Embase, PsycInfo, Web of Science, Cochrane Central Register of Controlled Trials, and the Health Technology Assessment databases in March 2016. Due to the nature of the review, results were limited to those published after 2000. No restrictions were imposed on language of publication. Searches were updated on September 12, 2017. Reference lists of relevant publications were screened to identify papers not picked up by the electronic searches. In addition, citations of selected key papers were searched.

Search Strategy

A detailed example of the search strategy used for Medline is outlined in Textbox 1. This search strategy was adapted for each of the databases.

Study Selection

For initial screening, a decision for inclusion was made based on title and where available, abstract. This was carried out by one researcher (LW) only, and for this reason, a cautious approach erring on the side of over-inclusion was used. Following this, 2 researchers (LW and KA) independently assessed all remaining papers for relevance. Disagreements were resolved by consensus after referring to the protocol. All discussions and decision making were documented. Where there was insufficient information to make a decision, authors were contacted for further information. If no response was received within 2 weeks, a final decision was made based on available information.

Table 1. PICOS (Population, Intervention, Comparator, Outcomes, Study) criteria.

Category	Criteria
Population	Male and female adults >18, no upper age limit, worldwide with any cancer diagnosis, receiving cancer treatment OR within 3 months of completing treatment. The cancer treatment to include any treatment with significant side effects (eg, systemic therapies, radiotherapy, biological therapies).
Intervention	Online systems for patients to report or manage symptoms and side effects during cancer treatment from home; Internet-based or -enabled systems, including mobile apps. Other forms of interactive health communication applications, eg DVDs, games were excluded. Purely educational systems not interactive in any way were excluded. Systems developed to assess and monitor purely psychosocial symptoms were excluded (eg, depression, anxiety, emotional coping or stress). Sleep and fatigue were included. Systems designed to be accessed at one time point only were excluded; access to the system had to be ongoing.
Comparator	Stage 2 only: The review included studies with any comparator (eg, randomized or nonrandomized studies), in addition to studies with no comparator (eg, feasibility studies).
Outcomes	Stage 1: Dependent on the nature and number of papers found, we aimed to characterize systems. For example, we identified if studies included features such as Monitoring of symptoms by health care professionals (HCPs), Alerts for severe symptoms sent to HCPs, Monitoring of symptoms by patients (eg, graphical or tabular), Automated feedback/advice based on responses, Access to symptom information, Communication with other cancer patients, Direct communication with HCPs (distinct from symptom monitoring by HCPs). Stage 2: We aimed to collect where available, information on engagement with systems and information on any patient-centered outcomes, including but not restricted to any QoL measures; self-efficacy measures including patient activation, patient empowerment, mastery; and patient satisfaction.
Study design	Stage 2 only: The review was not restricted to randomized controlled trials, and feasibility studies with any evaluation data were included. Patients had to be using the system over time, and there had to be at least one intended time point of use more than 3 weeks after baseline. This timeframe was selected as many standard chemotherapy treatments are administered every 3 weeks.

Textbox 1. Example of search strategy used (Ovid Medline).

1. Neoplasms/
2. oncolog*.mp.
3. cancer patient*.mp.
4. 1 or 2 or 3
5. Medical Informatics/
6. Telemedicine/
7. Mobile Applications/
8. Smartphone/
9. Self Report/
10. Self Care/
11. Self-Assessment/
12. (electronic adj2 (Patient report* or Patient-report* or Self report* or Self-report* or Self manage* or Self-manage* or Self monitor* or Self-monitor* or Symptom report* or Symptom-report* or Symptom manage* or Symptom-manage*)).mp.
13. (online adj2 (Patient report* or Patient-report* or Self report* or Self-report* or Self manage* or Self-manage* or Self monitor* or Self-monitor* or Symptom report* or Symptom-report* or Symptom manage* or Symptom-manage*)).mp.
14. (web* adj2 (Patient report* or Patient-report* or Self report* or Self-report* or Self manage* or Self-manage* or Self monitor* or Self-monitor* or Symptom report* or Symptom-report* or Symptom manage* or Symptom-manage*)).mp.
15. (remote* adj2 (Patient report* or Patient-report* or Self report* or Self-report* or Self manage* or Self-manage* or Self monitor* or Self-monitor* or Symptom report* or Symptom-report* or Symptom manage* or Symptom-manage*)).mp.
16. 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15
17. 4 and 16

Limit 17 to (humans and yr="2000 -Current")

Data Items

For Stage 1, basic data were extracted on authors, title, year of publication, and country of origin, in addition to the name (if any given) and type of system being described (eg, Web-based or mobile app). If the system did not already have a descriptive name, an arbitrary name was assigned (eg, System A). A preliminary list of common features was created based on existing knowledge and further developed throughout data extraction until a comprehensive list of common or important features was identified. Data were extracted from each publication on the presence of each feature. This was coded as "Yes" only if it was explicitly described in the publication, otherwise it was coded as a "#" For abstracts, if it was unclear whether or not a feature was present by information available in an abstract, this was classed as "Unable to determine." Where information was lacking, authors were not contacted for information. However, searches were undertaken for other publications related to the same system.

For Stage 2, data were extracted from studies with some form of system evaluation (eg, patient use of system or evaluation of efficacy). This included data on the number of patient participants, baseline demographics, disease and treatment type, duration of the evaluation, methods used to assess engagement, and actual usage or adherence. Where available, data were also extracted on any patient-centered outcomes used and results of evaluation.

Data Extraction

Data were extracted using the online Systematic Review Data Repository [38]. The form was piloted on 10 randomly selected papers and further refined. For Stage 1, three additional researchers (KA, BC, MA) each double-coded a number of allocated publications, totaling 36% (27/77) of the overall included publications. A high level of agreement (86%) was found. Discrepancies were resolved by referring back to the protocol and additional publications where available. For Stage 2, the same 3 researchers again each double-coded a proportion of the included publications totaling 46% (13/29) and 100% agreement was found.

Quality Assessment

Quality was assessed using the Downs and Black checklist for nonrandomized studies [39] and was undertaken alongside data extraction. It was deemed appropriate to assess only studies that included some feasibility/evaluation data, that is, publications included in Stage 2. Studies were given a score along a possible range of 0-26.

Synthesis of Results

A narrative synthesis was undertaken using the guidelines outlined by the Economic and Social Research Council [40]. Microsoft Excel was used to manage data. For Stage 1, information from multiple publications relating to the same systems was pooled to form a description of features. Where information was conflicting due to earlier and later iterations, the most recent description was used. For Stage 2, information

was collected on how patient engagement was assessed for any feasibility study or trial that included these data. For trial studies, information was collected on primary and secondary study outcomes and any results recorded. We then summarized these data to explore any relationships with system features identified in Stage 1.

Results

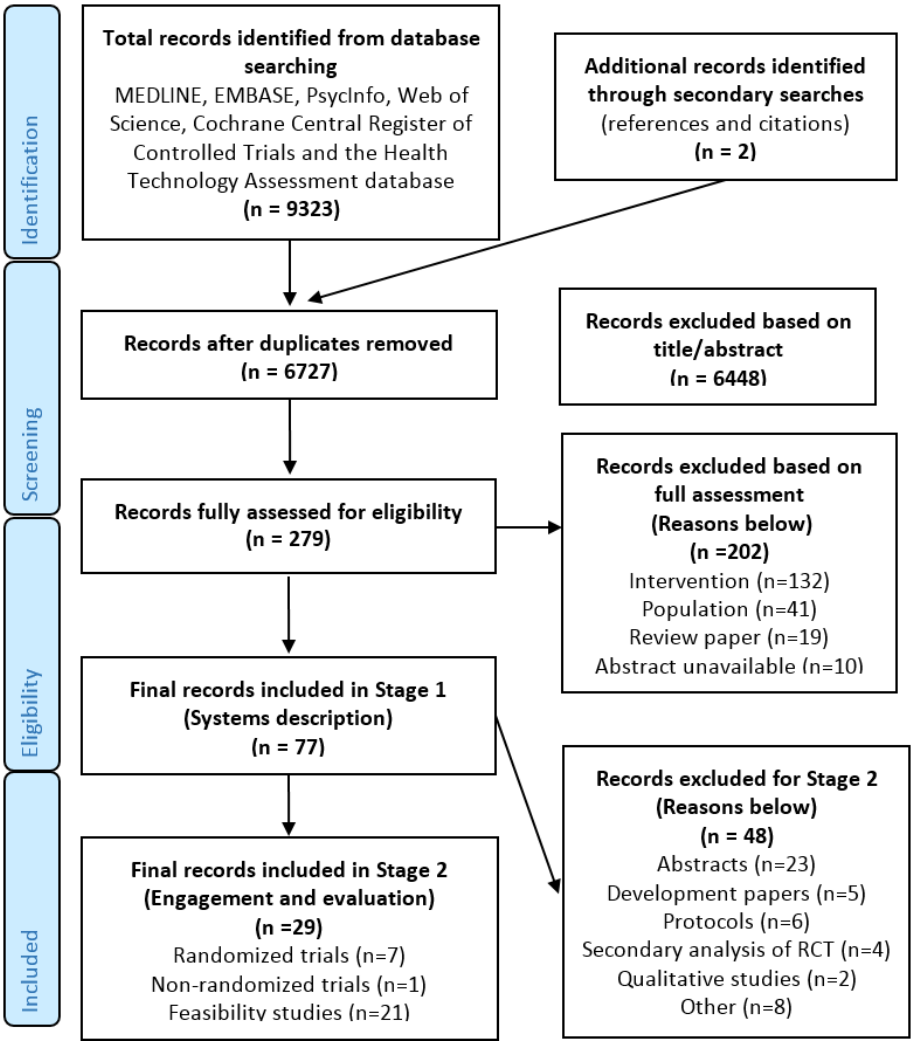
Study Selection

An overview of search and selection procedures is outlined in Figure 1. A total of 6727 publications were identified after removal of duplicate publications, including two publications identified from secondary searches (ie, citation and reference lists). All publications were in English. We assessed 279 publications for eligibility, and a total of 202 papers were excluded at this point based on predefined eligibility criteria (intervention, eg, not home-based or Web-based, n=132; population, eg, patients not on active treatment, n=41; discussion paper or systematic review, n=19; or abstract unavailable, n=10). We included 77 publications in Stage 1 of the review (ie, systems descriptions). A large proportion (23/77, 30%) of these publications were abstracts. The reasons for exclusions are outlined in Figure 1. Those 8 publications categorized under "Other" included 2 summary papers giving an overview of development and evidence for a system, a description of standard usability testing, a cost-effectiveness analysis, a content analysis of email communication within a system, a discussion of design approaches and methodology, an evaluation focusing on blood monitoring, and one publication where we were not able to access the full paper and did not receive a response from the authors when this was requested. We identified 29 publications for inclusion in Stage 2 of the review (ie, patient engagement and evaluation of systems). These were 21 feasibility studies and 8 controlled trials (7 randomized and 1 nonrandomized).

Stage 1: Description of Systems and Features

The 77 publications referred to 41 individual systems. Most originated from the United States (19/41, 46%) or the United Kingdom (6/41, 15%), and all publications were available in English. Systems were commonly Web-based (24/41, 56%), 27% (11/41) were mobile apps, 2 were both mobile and Web-based (2/41, 5%), and 22% (9/41) were Web-enabled mobile devices purposely designed for symptom reporting and were provided to patients for the duration of the study.

Figure 1. Summary of papers identified and subsequently excluded/included in this review.



Seven common system features were identified. Figure 2 outlines each of the features and its prevalence in the 41 identified systems. Features could be categorized broadly as supporting patients to monitor and manage their own symptoms, and to communicate with health professionals and one another, or supporting clinicians to monitor and manage patient symptoms.

Table 2 [4-8,11-20,41-102] provides an overview of each identified system and its associated publications, in addition to the presence or absence of each of the features identified in Figure 2.

Stage 2: Patient Engagement and Evaluation

Quality Assessment

Along a possible range of 0-26, the overall median quality assessment score of studies using the Downs and Black checklist was 17.0 (mean 16.2, SD 5.3, range 2-24). For trials described in the section on patient-centered outcomes [5,6,8,49,60,79,88,100], the median score was higher at 20.0 (mean 20.4, SD 2.6, range 17-24).

Patient Engagement

Table 3 [5,6,8,11-15,42,43,49,60,63,65,68,73-75,79,81,82,84,87,88,90,92,93,100,101] summarizes data on patient engagement from the 29 included studies (ie, 21 feasibility studies, 7 RCTs, and one non-RCT [88]). All 21 feasibility studies (100%) reported some data on patient engagement, although there was variation in how engagement was defined and measured. Three of the eight trials (38%) did not report any data on patient engagement [6,79,100].

Of the 29 studies, the most common method of assessing engagement was the number of symptom report completions or number of times the system was accessed (12/29, 41%) [15,49,60,63,65,68,74,87,88,90,92]. This was given as an overall figure for the whole sample [15,49,68,90,92], as an average per patient [13,15,65,74,90], or with a breakdown of the variance [63,87]. Nine studies (9/29, 31%) assessed adherence by number of actual completions/accesses in comparison to the number of expected completions/accesses [5,13,14,73,75,81,84,93,101]. This was reported as median or mean adherence of the overall sample for the duration of the study period [2,73,75,81,93,101], or with a breakdown of adherence at different time points [14,84]. Only 2 studies studies (2/29, 7%) categorized patients as users or nonusers dependent on predefined criteria [11,12]. Four studies (4/29, 14%) combined results of patients reporting from home and in clinic [11,13-15]. Not all studies reported on actual usage, and some used evaluation questionnaires with or without semistructured interviews to assess acceptability to patients [42,43,65,82].

Due to the variation in the methods of reporting, it was not possible to determine if there was any overall association between engagement and specific system features.

Figure 2. Overall summary of prevalence of identified system features.

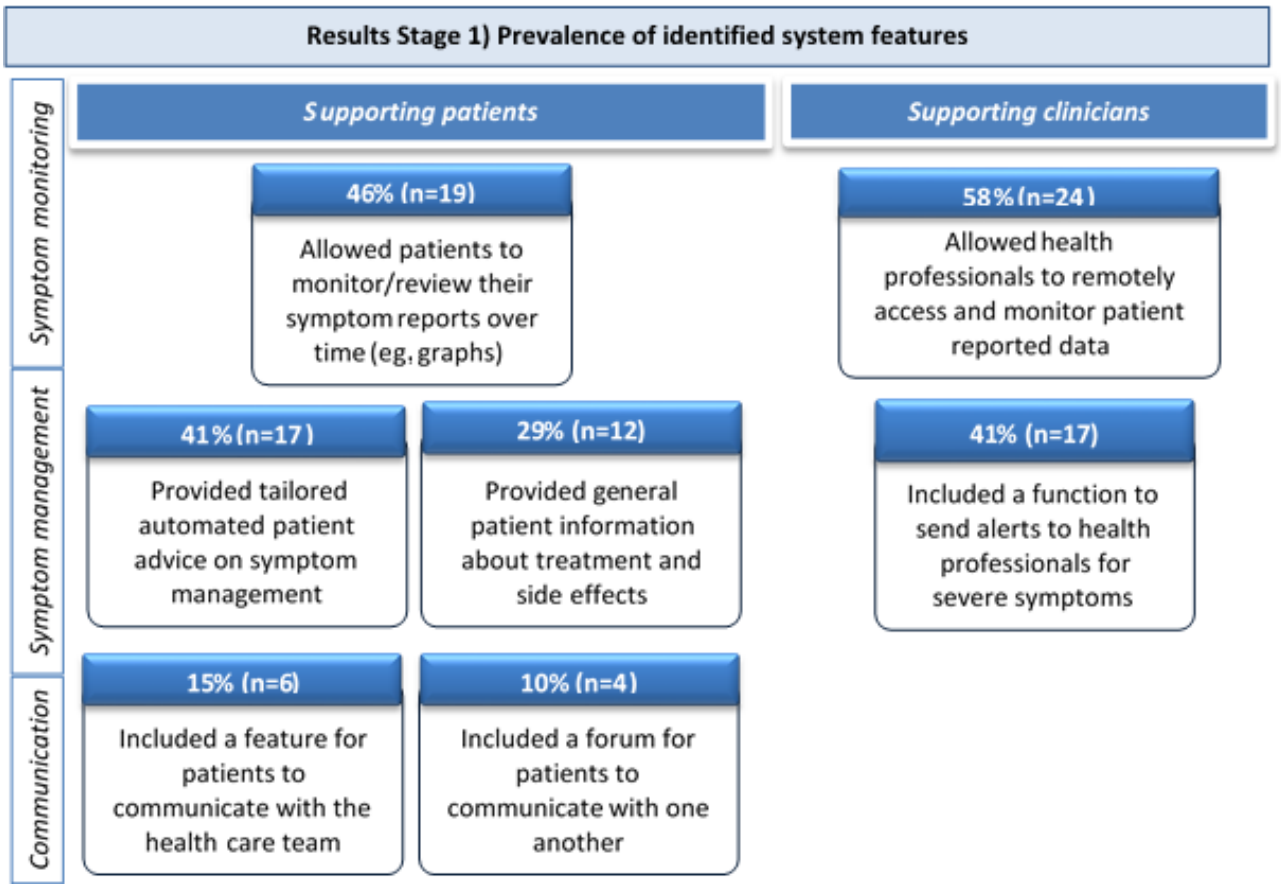


Table 2. Identified systems with description of features and associated publications[a].

System name (country)and type	Publication type (with relevant references)	Allowed health professional to remotely access and monitor patient reported data	Allowed patients to monitor their symptom reports over time (eg, graphs)	Included a function to send alerts to health professional for severe symptoms	Provided tailored automated patient advice on managing symptoms	Provided general patient info about cancer treatment and side effects	Included a feature for patients to communicate with the health care team	Included a forum for patients to communicate with one another
ASyMs (UK) Mobile device	Randomized trial [6], Secondary analysis of RCT[b] [41], Feasibility studies [42,43], Abstracts [44-47], Other [48]	✓	x	✓	✓	x	x	x
CASSY (USA) Web-based	Randomized trial [49]	x	✓	x	x	✓	x	✓
CHES (Austria) Web-based	Abstract [50]	–	–	–	–	–	–	–
COPE-CIPN (USA) Web-based	Other [51]	–	–	–	–	–	–	–
CORA (USA) Mobile app	Development paper [52], Protocol [53]	x	✓	x	✓	✓	x	x
eRAPID (UK) Web-based	Protocol [4], Abstracts [54-59]	✓	✓	✓	✓	✓	x	x
eSMART (UK) Mobile device	Protocol [7]	✓	✓	✓	✓	✓	x	x
ESRA-C (USA) Web-based	Randomized trial [60], Secondary analysis of RCT [61], Qualitative paper [62]	x	✓	x	✓	x	x	x
Healthweaver (USA) Web-based & mobile app	Feasibility study [63], Development paper [64]	x	✓	x	x	✓	x	x
HSM (UK) Mobile device	Feasibility study [65]	✓	x	✓	✓	✓	x	x
ICT-FP7 (France) Mobile device	Abstract [66]	✓	–	–	–	–	–	–
INTERAKTOR (Sweden) Web-based & Mobile app	Protocol [67]	✓	✓	✓	✓	✓	x	x

System name (country)and type	Publication type (with relevant references)	Allowed health professional to remotely access and monitor patient reported data	Allowed patients to monitor their symptom reports over time (eg, graphs)	Included a function to send alerts to health professional for severe symptoms	Provided tailored automated patient advice on managing symptoms	Provided general patient info about cancer treatment and side effects	Included a feature for patients to communicate with the health care team	Included a forum for patients to communicate with one another
KAIKU (Finland) Web-based	Feasibility study [68]	✓	x	x	x	x	✓	x
MADELINE (USA) Mobile app	Abstract [69]	–	–	–	–	–	–	✓>
MSKCC Web-Core (USA) Web-based	Abstract [70]	–	–	–	–	–	–	–
Onco-TREC (Italy) Mobile app	Development paper [71], Protocol [72]	✓	✓	✓	✓	x	✓	x
PatientViewpoint (USA) Web-based	Feasibility study [73]	✓	✓	✓	x	x	x	x
PaTOS (USA) Web-based	Feasibility study [74]	✓	x	x	x	x	x	x
Pit-a-pit (Korea) Mobile app	Feasibility study [75]	✓	x	x	x	x	x	x
PRISMS (Australia) Mobile device	Protocol [76], Abstract [77]	✓	✓	✓	✓	✓	x	x
PROCDIM (USA) Web-based	Abstract [78]	✓	✓	–	–	–	–	–
QoC Health Inc (Canada) Mobile app	Randomized trial [79], Other [80]	✓	x	✓	x	x	x	x
RemeCoach (Belgium) Mobile device	Feasibility study [81]	x	x	✓	x	x	x	x
SCMS (Singapore) Web-based	Feasibility study [82], Other [83]	✓	x	x	x	✓	✓	x
STAR (USA) Web-based	Randomized trial [5], Feasibility studies [11-15,84]	x	✓	✓	✓	x	x	x
The Health Buddy (R) (USA) Mobile device	Development paper [85]	✓	x	✓	✓	x	x	x

System name (country)and type	Publication type (with relevant references)	Allowed health professional to remotely access and monitor patient reported data	Allowed patients to monitor their symptom reports over time (eg, graphs)	Included a function to send alerts to health professional for severe symptoms	Provided tailored automated patient advice on managing symptoms	Provided general patient info about cancer treatment and side effects	Included a feature for patients to communicate with the health care team	Included a forum for patients to communicate with one another
WebChoice (Norway) Web-based	Randomized trial [8], Secondary analysis of RCT [17,18], Qualitative paper [20], Other [16,19]	x	✓	x	✓	✓	✓	✓
WRITE (USA) Web-based	Abstract [86]	✓	–	–	✓	–	–	–
System A (USA) Web-based	Feasibility study [87]	x	x	✓	x	x	x	x
System B (The Netherlands) Web-based	Nonrandomized trial [88], Development paper [89], Feasibility study [90]	✓	✓	✓	x	✓	✓	✓
System C (USA) Web-based	Other [91]	–	–	–	–	–	–	–
System D (Sweden) Mobile app	Feasibility study [92]	✓	✓	✓	✓	✓	x	x
System E (UK) Mobile device	Feasibility study [93]	✓	✓	✓	✓	x	x	x
System F (Canada) Web-based	Abstract [94,95]	–	✓	–	✓	–	✓	–
System G (Denmark) Web-based	Abstract [96]	–	✓	–	✓	–	–	–
System H (UK) Mobile device	Other [97]	✓	x	✓	x	x	x	x
System I (USA) Web-based	Abstract [98]	–	–	–	–	–	–	–
System J (USA) Web-based	Abstract [99]	✓	–	–	–	–	–	–
System K (Switzerland) Mobile app	Randomized trial [100]	✓	✓	x	x	x	x	x
System L (USA) Mobile app	Feasibility study [101]	✓	x	x	x	x	x	x

System name (country)and type	Publication type (with relevant references)	Allowed health professional to remotely access and monitor patient reported data	Allowed patients to monitor their symptom reports over time (eg, graphs)	Included a function to send alerts to health professional for severe symptoms	Provided tailored automated patient advice on managing symptoms	Provided general patient info about cancer treatment and side effects	Included a feature for patients to communicate with the health care team	Included a forum for patients to communicate with one another
System M (USA) Mobile app	Abstract [102]	–	–	–	–	–	–	–

[a]"✓" denotes feature is present, "x" denotes feature is not present, and "–" denotes that it was not possible to determine whether feature was present or not.

[b]RCT: randomized controlled trial.

Patient-Centered Outcomes

All the trials used some measure of patient-centered outcome to evaluate system efficacy, most commonly validated QoL and symptom and psychosocial outcome measures. Table 4 outlines each trial [5,6,8,49,60,79,88,100], the intervention and comparator groups, outcomes reported, and a summary of the results.

Global Quality of Life

CASSY [49] and STAR [5] interventions both demonstrated improvements in overall QoL. However, in addition to the online component, CASSY included access to a collaborative care coordinator with experience in cognitive behavioral therapy and psycho-oncology, which is likely to have contributed to the efficacy. In the STAR study, patients were allocated to computer-experienced and inexperienced groups prior to randomization and only the computer-experienced group had access to the system from home. Results are pooled, making it difficult to assess efficacy for our purposes. No significant impact on QoL was found for WebChoice [8].

Physical Symptoms

An overall reduction of symptom distress was found in the studies assessing Electronic Self-Report Assessment-Cancer (ESRA-C) [60] and WebChoice [8]. However, in addition to the online intervention, ESRA-C also included a communication coaching component to improve symptom disclosure to physicians. System B [88] was found to have significant positive impact on the general physical complaints subscale compared to the control group.

Advanced Symptom Management System (ASyMs) [6] and Comprehensive Electronic Cancer Support System for the Treatment of Cancer Related Symptoms (CaSSY) [49] both demonstrated positive impact on levels of fatigue while System K [100] demonstrated a lesser decline in functional activity in contrast to the control group, but this was not significant. Both ASyMs and System K were evaluated using the same measure as used to assess symptoms in the intervention, which may have affected results.

Self-Efficacy

WebChoice [8] and System B [88] both demonstrated a positive impact on self-efficacy. However, for System B, this was assessed only as a subscale of a main measure. System K [100] reported an improvement in patient empowerment; however, this was assessed using a single item regarding using the Internet for information seeking, which is unlikely to be a reliable measure.

Other Psychosocial Outcomes

CASSY [49] and WebChoice [8] demonstrated significant reductions in depression in intervention compared to control groups. System B [88] demonstrated no difference on the depression subscale of a QoL measure but a significant impact on state anxiety and fear related to specific head and neck problems. WebChoice demonstrated no impact on social support [8]. QoC Health Inc [79] was primarily assessed on number of hospital contacts but also included patient scores of convenience and satisfaction using a simple 5-point Likert scale and found an impact for convenience, but not for patient satisfaction.

Due to the considerable variation in outcomes used and study design, it was not possible to assess any relationships between outcomes and system features.

Table 3. Overview of patient engagement data.

System name, patient group (patients, N), treatment type and study duration, quality assessment score (QAS)	Method of evaluation/ patient engagement	Brief summary of findings
Feasibility studies (n=21)		
ASyMS-R [42], Lung (N=16) During and 1 month after thoracic radiotherapy QAS=19	Evaluation questionnaire and semistructured interviews	Actual usage not reported Patients perceived it to positively impact on care and promote timely reporting and management of symptoms
ASyMS [43], Colorectal or lung (N=18) During 2 cycles of chemo QAS=15	Evaluation questionnaire	Actual usage not reported Patients reported it helped monitor symptoms, promote self-care, and improve symptom management
HealthWeaver [63], Breast (N=9) Undergoing active treatment, 4 weeks QAS=8	# of completions/ accesses	All patients used website at least 3x/week, 7 patients used it almost daily Phone component used almost daily by 5 patients, 3x/week by 1 patient, and 1-2x/week by 3 patients
HSM [65], Lung or colorectal (N=18) During 2 cycles of chemo QAS=10	# of completions/ accesses and evaluation questionnaires	All patients completed 1-34 symptom reports, average 14 overall (SD 10.2) High variation in use of self-management advice Patients found system easier to use and more useful than expected
Kaiku [68], Head & neck (N=5) Radiotherapy, during and 1 month after QAS=12	# of completions/ accesses	514 symptoms reported (including zero grades) 23 questionnaires completed 38 messages sent
PatientViewpoint [73], Breast or prostate (N=47) Medical oncology treatment UTD - 3 onsite visits (not specified) QAS=15	# of accesses/ expected accesses	190/224 symptom reports completed (85%) Median expected questionnaires completed by individual patients was 71% Majority of questionnaires completed offsite (n=160; 87%)
PaTOS [74], Any disease site (N=30) Chemo, 10 weeks QAS=6	# of completions/ accesses	28/30 patients observed for 10 weeks Total 231 accesses, 193 fully completed Total of 1870 symptoms observations (average 69 per patient, 1.5 per day)
Pit-a-pit [75], Breast (N=30) Neo-adjuvant chemo, 90 days QAS=14	# of accesses/ expected accesses	1215/2700 responses (compliance=45.0 %) Median patient-level reporting rate was 41.1% (range 6.7-95.6%)
RemeCoach [81], Advanced solid tumors, eg, colorectal, gastric-esophageal, and pancreatic adenocarcinoma (N=11) Duration of Teysuno treatment QAS=18	# of accesses/ expected accesses	Average daily compliance 91.2% Could not determine longitudinal compliance because of the low patient number using the coach for an acceptable duration of time
SCMS [82], Breast, lung, or colorectal (N=4) During 4 cycles of chemo QAS=10	Evaluation questionnaire	All patients completed at least 1 symptom report Questionnaire revealed patients found system useful and easy to use
STAR [84], Gynecologic malignancy (N=49) Laparotomy, 6 weeks QAS=20	# of accesses/ expected accesses	Compliance of patients gradually decreased 92% of patients completed preoperative session, and 74% completed Week 6 session Majority of patients (82%) completed at least 4/7 total sessions in STAR

System name, patient group (patients, N), treatment type and study duration, quality assessment score (QAS)	Method of evaluation/ patient engagement	Brief summary of findings
STAR [11], Gynecologic malignancy (N=80) Chemo, 8 weeks QAS=16	Users/nonusers (logged in/did not log in)	Patients could access from home or in clinic 25% used only in clinic waiting area, remainder logged in from home and clinic Most patients with home computers (83%) logged in from home without reminders
STAR [12], Not specified (N=180) Chemo, 8 weeks QAS=8	Users/nonusers (logged in/did not log in)	Patients could access from home or clinic 2/3 voluntarily logged in from home computers without prompting
STAR [13], Thoracic malignancies (N=107) Chemo, 16 months QAS=20	# of accesses/ expected accesses	Patients could access from home or clinic 16 patients (15%) accessed system from home Home users accessed system more frequently than those using in clinic (avg=23 sessions, range 3-144) vs (avg=9, range 1-36) respectively
STAR [15], Lung, gynecologic, breast, genitourinary (N=286) Duration of chemo QAS=19	# of completions/ accesses	Patients could access from home or in clinic Total 8690 logins (median 17 logins per patient), avg 0.9 logins per patient per week 71% from home and 29% from clinic
STAR [14], Gynecologic malignancy (N=96) Laparotomy, preoperatively & weekly 6-wks postlaparotomy QAS=17	# of accesses/ expected accesses	74% (n=71) completed at least 4/7 surveys and were considered responders 63% (n=69) completed preoperative session. Remaining completed subsequent surveys. 9 (9%) patients completed only 1 survey
System A [87], Hepatobiliary and GI (N=20) Preoperatively and 2 weeks after discharge for curative resection QAS=17	# of completions/ accesses	65% (13/20) completed 8 symptom assessments 75% (15/20) completed 4 QoL assessments Mean 7 minutes to complete MD Anderson Symptom Inventory and mean 4 minutes to complete EuroQoL-5D-5L
System B [90], Head and neck cancer (N=36) Surgery, 6 weeks QAS=17	# of completions/ accesses	All patients used system (total sessions=982) Avg no of sessions was 27.3 (SD 18.4, range 4-69) Avg session 12 minutes, longest session 1 hour 38 minutes
System D [92], Prostate (N=9) Radiation, 2 weeks QAS=13	# of completions/ accesses	Patients reported for mean of 10 days Estimated time for report 5 minutes Self-care advice accessed by 85%, who logged 20 views at 34 symptoms 59 alerts: 55 yellow and 4 red
System E [93], Colon (N=6) Complete resection, during 2 cycles of chemo QAS=11	# of accesses / expected accesses	Data entry compliance was excellent (98% of the twice-daily input was complete) from all 6 patients with the exception of one question
System L [101], Head and neck (N=22) Duration of radiation (approx. 5-7 weeks) QAS=16	# of accesses/ expected accesses	Median compliance 71% (interquartile range [IQR], 45%-80%) 6 patients (27%) compliance ≥80%, 2 patients (9%) 100% compliant Median reports submitted 34 (IQR 21-53)
Randomized controlled trials (RCTs; n=7; n refers to # of patients expected to use the system [ie, intervention arm])		
ASyMS [6], Breast, lung, or colorectal (N=56) 4 cycles of chemo QAS=22	Not reported	Not reported

System name, patient group (patients, N), treatment type and study duration, quality assessment score (QAS)	Method of evaluation/ patient engagement	Brief summary of findings
CASSY [49], Any cancer diagnosis (N=144) Chemo, radiation, or surgery, 6 months QAS=19	# of completions/ accesses	Total number of page views=1491 Total duration in minutes=1813.9 Total views and duration given for individual patients
ESRA-C [60], Diagnosis of cancer (N=374) Any therapeutic regimen, UTD, over 4 visits QAS=24	# of completions/ accesses	Median access rate of 4 (range 2-4) at study time points Median access rates of 1 (range 0-8) at voluntary times
QoC Health Inc [79], Breast (N=32) Reconstructive surgery, 30 days QAS=23	Not reported	Not reported
STAR [5], Metastatic breast, genitourinary, gynecologic, or lung (N=286) Duration of chemo QAS=20	# of accesses/ expected accesses	Computer experienced (home access) and inexperienced (clinic access) figures combined Avg 73% completed a self-report at any given clinic visit (includes clinic completions)
WebChoice [8], Breast or prostate (N=162) Surgery plus radiation, chemo, hormone therapy, or a combination, 1 year QAS=23	# of completions/ accesses	77% logged on at least once 23% never logged on Of 103 (64%) who logged on more than once, avg logons=60 times (range 2-892)
System K [100], Breast cancer (N=95) Adjuvant or neo-adjuvant chemo, 6 weeks QAS=18	Not reported	Not reported
Non-RCT (n=1) (n refers to # of patients expected to use the system [ie, intervention arm])		
System B [88], Head and neck cancer (N=39) Surgery, 6 weeks QAS=17	# of completions/ accesses	Avg # of sessions=27, avg length of session=12 mins Avg # of completions=12.6 Avg # of messages=4.5

Table 4. Overview of patient-centered outcomes data.

Study, population (N); study design	Intervention and comparator groups	Outcomes reported	Summary of results
ASyMs [6], Breast, lung, or colorectal (N=112); 2-arm randomized controlled trial (RCT), 4 cycles of chemo	Intervention (N=56): Asked to complete a symptom questionnaire integrating Common Toxicity Criteria Adverse Events (CTCAE) grading system and Chemotherapy Symptom Assessment Scale Symptom information sent in real time to the study server Patients receive severity dependent tailored self-care advice on mobile phone interface Evidence-based risk assessment tool alerts clinicians via a dedicated 24-h pager system of any severe symptoms Comparator (N=56): Standard care following local guidelines and procedures related to monitoring and reporting of chemo-related toxicity including written and verbal information from nurses administering chemo	Primary outcomes: Paper version of online questionnaire; Comparison between groups on mean scores from 4 paper-based completions at baseline and before each chemo cycle	Higher reports of fatigue (*P*=.04) and lower reports of hand-foot syndrome (*P*=.03) in control group compared with intervention group No difference on nausea, vomiting, diarrhea, or sore mouth/throat
CASSY [49], Any diagnosis of cancer Chemo, radiation, or surgery (N=261) 2-arm RCT, 6 months	Intervention (N=144): Access to psycho-educational website where patients could record and monitor symptoms via graphs and journal Access chat room to communicate with other study patients Audiovisual and resource library including relaxation techniques and educational videos Phone contact (approx. every 2 weeks) with a collaborative care coordinator with training and experience with cognitive-behavioral therapy and psycho-oncology Comparator (N=117): Usual care provided by medical team plus assessment of symptoms and blood draws at the same time as intervention patients to evaluate efficacy of intervention	Primary outcomes: Depression (Centre for Epidemiologic Studies-Depression≥16), Pain Brief Pain Inventory, Anemia (Functional Assessment of Cancer Therapy [FACT]-Anemia), Hepatobiliary (FACT-Hep) Secondary outcomes: Serum cytokines levels and Natural Killer Cell (NK), Comparison at 6 months follow-up	Reductions of fatigue at 6 months (*P*=.09) Statistically and clinically significant changes in overall QoL (*P*=.05) Reductions in pain and depression Medium effect size for NK cell number (Phi=0.491) at 6 months (chi-square=3.62, *P*=.057)
ESRA-C [60], Diagnosis of cancer Any therapeutic regimen (N=779) 2-arm RCT, UTD, over 4 visits	Intervention (N=374): Participants completed cancer symptoms and QoL (SxQoL) assessments at each study time point and ad lib between visits Summary reports delivered to clinicians Self-management advice given for 3 symptoms Coaching to verbalize issues to health care team Alert to contact health care team for severe symptoms Patients could monitor symptoms via graphs and journal Self-care strategies and coaching available at any time Comparator (N=378): Participants completed assessments at each study time point Summary reports delivered to clinicians Research staff verbally notified health care team of any severe symptoms reported at clinic visit Both groups were provided the same patient education typically available in each clinic	Primary outcomes: Symptom Distress Scale (SDS) plus 2 items (impact on sexual activity and interest, fever/chills) to form SDS-15, End point was change in SDS-15 total score from baseline to the end-of-study time point	Intervention had lower symptom distress; mean change in SDS-15 score was 1.27 ([SD], 6.7) in control (higher distress) and -0.04 (SD 5.8) in intervention (lower distress) SDS-15 score reduced by estimated 1.21 (95% CI 0.23-2.20; *P*=.02) in intervention vs control group

Study, population (N); study design	Intervention and comparator groups	Outcomes reported	Summary of results
QoC Health Inc [79], Breast cancer Surgery (N=65) 2-arm RCT, 30 days	Intervention (N=32): Follow-up visits at 1 and 4 weeks replaced with examination of surgical site via photos submitted through mobile app, plus completion of pain visual analog scale and quality of recovery 9-item questionnaire Reporting began after discharge from recovery room Email reminder if submission not received Surgeon used wireless interface to access data and monitor patient's condition Severe scores flagged in the database for quick viewing. Red flags prompted in-person follow-up Physicians summarized data from mobile app using prototypical subjective, objective, assessment, and plan note at 1 or more time points during 30-day monitoring period Comparator (N=33): Patients in conventional follow-up group had planned clinic follow-up at approx. 1 week and 4 weeks after operation	Primary outcomes: Total number of follow-up visits (including specialists, family physician, and emergency department), Total number of phone calls and emails to health care team, Satisfaction and convenience scores using 5-point Likert scale, Postop complications	Control group more likely to attend in-person follow-up care first 30 days after surgery (95% CI 0.24-0.66; *P*<.001) Intervention group sent more emails than control group (IRR 4.13; 95% CI 1.55-10.99; *P*=.005) Intervention group reported higher convenience scores (IRR 1.39; 95% CI 1.09-1.77; *P*=.008)
STAR [5], Metastatic breast, genitourinary, gynecologic, or lung cancers (N=766) Before randomization, participants assigned to subgroups (computer-experienced and computer-inexperienced) Only computer-experienced intervention used system from home Duration of chemo	Intervention (N=286): Remote access to Web-based interface including questions adapted for patient use from CTCAE Triggered email alerts to nurses when patient-reported symptom worsened by 2 points or reached an absolute grade Report tracking participant's symptoms printed at each clinic visit for both nurse and treating oncologist No specific guidance provided to clinicians on actions to take in response to alerts or printed symptom profiles Comparators: Intervention – Computer-inexperienced (N=155): Similar to main intervention group but accessed system in clinic only and did not have remote access Computer-experienced – Usual care (N=253) Computer-inexperienced – Usual care (N=72): Usual care for the computer-experienced and computer-inexperienced subgroups consisted of standard procedure for monitoring and documenting symptoms Symptoms discussed and documented in the medical record during clinical encounters between patients and oncologists Patients encouraged to initiate phone contact between visits for concerning symptoms	Primary outcomes: EuroQol EQ-5D Index given via paper at clinic visits every 12 ± 4 weeks throughout study Secondary outcomes: Survival at 1 year, Time to first emergency room visit and time to first hospitalization, Time receiving active cancer treatment, Number of nursing calls to patients	Combined results for computer-experienced (home system) & computer-inexperienced (clinic only) intervention Greater improvement in Health-Related QoL scores in intervention vs usual care arm (34% vs 18%) and worsened among fewer (38% vs 53%; *P*<.001) Greater survival in intervention arm (69% vs 75%, *P*=.05) Fewer emergency room visits in intervention (34% vs 41%, *P*=.02) Intervention received chemo for longer (8.2 vs 6.3 months, *P*=.002) No difference in number of nursing calls to patients

Study, population (N); study design	Intervention and comparator groups	Outcomes reported	Summary of results
WebChoice [8], Breast or prostate cancer Surgery plus additional treatment of either radiation, chemo, hormone therapy, or a combination of those) (N=325) 2-arm RCT, 1 year	Intervention (N=162): Assessment component to monitor and report symptoms, problems, and priorities for support along physical, functional, and psychosocial dimensions Patients receive automated tailored self-management advice based on responses Patients receive advice to contact health care team when appropriate Info can be used to create a self-care plan Info section with access to other reliable, relevant Web resources Communication section including (1) unrestricted support forum for group discussion, allowing patients to post messages anonymously, (2) question-and-answer area where patients can privately ask questions of expert nurses in cancer care Access to diary to keep personal notes Comparator (N=163): In addition to a letter giving their group assignment, participants receive info sheet with suggestions for publicly available, cancer-relevant websites	Primary outcomes: Memorial Symptom Assessment Scale Short Form Secondary outcomes: Center for Epidemiological Studies Depression scale, Cancer Behaviour Inventory, 15D Health-related QoL, Medical Outcome Study Social Support Survey	Between-group differences significant for the Global Distress Index only (t=4.42; P=.037) No significant differences on the other subscales or total score or any secondary outcomes Experimental group showed significant improvements in depression (t=-2.71; P=.007) Control group had worsened self-efficacy (t=-2.82; P=.005) and Health-related QoL scores significantly (t=-2.77; P=.006),
System B, Van den Brink [88] Head and neck cancer Surgery (N=163) Nonrandomized trial, 6 weeks	Intervention (N=39): Provided with a laptop Patients could be monitored at home (by means of electronic questionnaires) Could communicate (send messages) to team Access to information Communicate with fellow sufferers (via a forum) Comparator (N=128): Routine follow-up apps at 2 and 6 weeks after discharge Patients could contact their care providers, both in- and outside hospital, if considered necessary	Primary outcomes: QoL measure assessed state anxiety, object anxiety, feelings of depression, uncertainty, feelings of insecurity, loss of control, self-efficacy, loneliness, and complaints	Intervention had significantly better change from baseline at 6 wks for state anxiety (P=.01), fear related to specific head and neck problems (P=.02), physical self-efficacy (P=.03), perceived abilities in swallowing and food intake (P=.04), general physical complaints (P=.02)
System K, Egbring [100], Breast cancer Adjuvant or neo-adjuvant chemo (N=139) 3 arm RCT, 6 weeks	Intervention (N=49): App and physician: Patients used mobile app and reviewed reported data with treating physician at scheduled visits Patients could report daily functional activity or symptoms with indication of severity Patients could edit a quick list of their preselected symptoms or select any of the 48 symptoms made available from the CTCAE listing Treating physician enabled access to review and discuss electronically reported symptoms during scheduled visits Comparators: Attention-control group (N=46) App only: Patients instructed to use the mobile app without physician review Control group (N=44): Received regular physician support	Primary outcomes: Daily functional activity measured by ECOG Secondary outcomes: Symptom reporting (intervention group and attention control group only), Patient-physician communication (measure not specified), Patient Empowerment (measure not specified)	Control groups showed greater decline in functional activity versus intervention but not significant At last visit, intervention & attention control patients reported fewer concentration issues than control group (P=.002) At third visit, significantly more intervention & attention control patients confirmed use of Internet for disease information compared vs control

Discussion

Principal Findings

The main aim of this review was to systematically describe and assess the features and functions of current systems available for patients to report and manage side effects of cancer treatment. We also wanted to focus on understanding the level of evidence indicating whether key system features are associated with better patient system engagement and patient outcomes.

In Stage 1 of the review, we identified a total of 41 individual systems. There was significant variation between systems,

though published descriptions of systems were often limited. We developed a taxonomy of features that were then classified into those supporting clinicians to deliver patient care in an innovative way and those aimed to support patients to better self-manage their condition and identify when medical input may be needed. This was successfully applied to describe the presence or absence of common system features.

The review of features highlighted some interesting findings. It was surprising to note that while over half (58%) of systems had the facility for health care providers to monitor patient data over time, fewer than half (46%) included the facility for patients to monitor and review their own data. Given the available evidence suggesting that self-monitoring is generally beneficial to support patients' self-management [28,33,103], this feature could be very important to improve efficacy of systems and in most cases, may be relatively easy to implement. Similarly, less than half of the systems (41%) included a feature for delivering advice to support patients to self-manage symptoms and less than a third provided patients with access to general educational information. The two least common features were facilities to support communication between patients and health care providers (15%) and communication between patients themselves, respectively (10%). Previous research has indicated that these features are highly valued and utilized by patients [20,22,29,33]. It is likely that these features are less common due to complexities in their implementation and maintenance. For example, it may be difficult to engage busy clinicians to respond to patient communication in this way, and there are ethical considerations around the need to moderate patient forums that are endorsed by a health care facility.

In Stage 2 of the review, we found little agreement on how patient engagement with systems was defined, measured, or reported, which meant it was not possible to compare levels of engagement across studies or make any conclusions on relationships with system features. Our review also indicated heterogeneity in terms of outcomes used to evaluate systems. Even of those that focused on symptoms or global QoL, the variation in methods and measures used made meaningful comparison impossible.

Due to the heterogeneous nature of reporting engagement and outcomes, we were unable to explore any relationships with system features. Our findings are similar to other reviews undertaken in this area, which have also found that poor assessment and reporting of patient engagement with systems makes comparison between studies difficult. Brower et al made quantifiable and comparable reports of engagement as part of their inclusion criteria for their review, and results indicated that facility for communication with other patients may be a very influential factor in patient engagement and needs careful consideration during system design [22]. However, other oncology specific reviews have found that methods of assessing and reporting patient engagement were too heterogeneous to make meaningful conclusions [104,105]. We identified only 8 trials (7 randomized and 1 nonrandomized) that evaluated systems, none of which reported any analysis on relationships between engagement and outcomes, and 3 of which did not report any data on patient engagement at all. This does not seem to be unique to oncology. Donkin et al [106] set out to review the impact of patient engagement with e-therapies across a range of disease groups and similarly found that this is not a link that is routinely explored.

Robust evidence supporting the value of systems for patient-centered outcomes was limited, with a large proportion of feasibility studies identified and even fewer RCTs. While all trials used some measure of patient-centered outcome to evaluate systems, a wide range of assessment tools were used, again making comparison difficult. In addition, 2 studies used the same measure for symptom assessment as part of the intervention, as for the outcome measure. Only 3 trials reported any measure of self-efficacy or patient empowerment, one of which used a study-specific nonvalidated measure [79], and another that was assessed using a subscale of a global QoL measure [88]. There is an array of evidence to suggest that online interventions can have a positive impact on self-efficacy and patient activation levels [30,32,33,107]. Growing evidence suggests that self-efficacy and patient activation play a significant role in symptom management and quality of life throughout cancer treatment [108,109] and are associated with an array of improved health behaviors and health outcomes [110-112] and lower use of hospital resources [113]. The reviewed systems generally demonstrate positive outcomes for patients as has been found in other reviews [31].

To our knowledge, this is the first systematic review in this field to identify and characterize all available systems for patients to report and manage side effects of cancer treatment, in addition to evidence on patient engagement and patient-centered outcomes.

Limitations

In order to meet the aims of the review, we included many publications that provided limited information about the system evaluated and some of which were of poor quality. However, we felt that this was necessary in order to meet the aims of the review and evaluate all evidence. Due to limitations on available resources, the initial stage of study selection (ie, assessment of titles and abstracts) was undertaken by a single reviewer. This is a limitation of our methodology and may have resulted in some bias of inclusion. To address this, a cautious approach erring on the side of over-inclusion was adopted, in order for records to be fully assessed by 2 researchers in the next stage of the review.

Due to the heterogeneous nature of study designs and methods of reporting engagement and outcomes, we were unable to explore any relationships with system features. This is a field of research that is still in its infancy, and the large number of feasibility studies and abstracts identified are likely to be indicative of this. The search was last updated September 2017. Due to the fast-moving nature of this field of research, it is likely that additional publications will be available by the time of publication. This is a common limitation of systematic reviews that is particularly pertinent with reviews of technology [114]. We did identify a number of protocols for planned quality trials that may contribute to a more in-depth understanding of associations between system features, adherence, and outcomes in the future [4,7,53,67,72,76]. In addition, we have not explored how issues with implementing systems into clinical practice

may have affected the efficacy of systems. A discussion of these issues is outside the scope of this review but has been well-documented elsewhere [115].

Conclusions

There is a real need for evidence-based guidance on developing, evaluating, and reporting systems. Based on this systematic review, we propose a taxonomy for characterizing system features to guide future development, improvement, and implementation of such systems. More work is needed to develop guidance for standardized reporting of patient engagement both in feasibility studies, and in evaluation trials. This is a complex and multifaceted issue, and it is important that barriers and facilitators to engagement are shared to help the evolution of more sustainable and valuable systems. Similarly, the development of guidance for the evaluation of systems is necessary. Variation in approaches to design and implementation will rightly affect outcomes chosen to evaluate efficacy [104,105]. However, there is enough commonality between systems to call for a set of recommended core outcomes to be developed [116]. More work is needed to develop this, and this is something we will work towards in the future. However, based on this review we recommend that all system evaluations include (1) a description of the system using our taxonomy of system features, (2) measures of feasibility and engagement, (3) patient-centered outcomes focusing on QoL and symptom improvement, in addition to those focusing on self-efficacy and patient activation, and (4) a measure of health economics. This will facilitate synthesis of evidence in order to improve the design of systems and make them practically useful for both patients and clinicians.

Conflicts of Interest

None declared.

References

1. Furstenberg CT, Carter JA, Henderson JV, Ahles TA. Formative evaluation of a multimedia program for patients about the side effects of cancer treatment. Patient Educ Couns 2002 May;47(1):57-62. [Medline: 12023101]
2. Warrington L, Holch P, Kenyon L, Hector C, Kozlowska K, Kenny AM, et al. An audit of acute oncology services: patient experiences of admission procedures and staff utilisation of a new telephone triage system. Support Care Cancer 2016 Dec;24(12):5041-5048. [doi: 10.1007/s00520-016-3370-4] [Medline: 27540738]
3. National Confidential Enquiry into Patient Outcome and Death. 2008. For better, for worse? A review of the care of patients who died within 30 days of receiving systemic anti-cancer therapy URL: https://www.ncepod.org.uk/2008report3/Downloads/SACT_report.pdf [accessed 2018-11-28] [WebCite Cache ID 74GLidF7k]
4. Absolom K, Holch P, Warrington L, Samy F, Hulme C, Hewison J, eRAPID Systemic Treatment Work Group. Electronic patient self-Reporting of Adverse-events: Patient Information and aDvice (eRAPID): a randomised controlled trial in systemic cancer treatment. BMC Cancer 2017 May 08;17(1):318 [FREE Full text] [doi: 10.1186/s12885-017-3303-8] [Medline: 28482877]
5. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
6. Kearney N, McCann L, Norrie J, Taylor L, Gray P, McGee-Lennon M, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 2009 Apr;17(4):437-444. [doi: 10.1007/s00520-008-0515-0] [Medline: 18953579]
7. Maguire R, Fox P, McCann L, Miaskowski C, Kotronoulas G, Miller M, et al. The eSMART study protocol: a randomised controlled trial to evaluate electronic symptom management using the advanced symptom management system (ASyMS) remote technology for patients with cancer. BMJ Open 2017 Dec 06;7(5):e015016 [FREE Full text] [doi: 10.1136/bmjopen-2016-015016] [Medline: 28592577]
8. Ruland CM, Andersen T, Jeneson A, Moore S, Grimsbø GH, Børøsund E, et al. Effects of an internet support system to assist cancer patients in reducing symptom distress: a randomized controlled trial. Cancer Nurs 2013;36(1):6-17. [doi: 10.1097/NCC.0b013e31824d90d4] [Medline: 22495503]
9. Yang LY, Manhas DS, Howard AF, Olson RA. Patient-reported outcome use in oncology: a systematic review of the impact on patient-clinician communication. Support Care Cancer 2018 Dec;26(1):41-60. [doi: 10.1007/s00520-017-3865-7] [Medline: 28849277]
10. Basch E. Patient-Reported Outcomes - Harnessing Patients' Voices to Improve Clinical Care. N Engl J Med 2017 Jan 12;376(2):105-108. [doi: 10.1056/NEJMp1611252] [Medline: 28076708]
11. Basch E, Artz D, Dulko D, Scher K, Sabbatini P, Hensley M, et al. Patient online self-reporting of toxicity symptoms during chemotherapy. J Clin Oncol 2005 May 20;23(15):3552-3561. [doi: 10.1200/JCO.2005.04.275] [Medline: 15908666]
12. Basch E, Artz D, Iasonos A, Speakman J, Shannon K, Lin K, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. J Am Med Inform Assoc 2007;14(3):264-268 [FREE Full text] [doi: 10.1197/jamia.M2177] [Medline: 17329732]

13. Basch E, Iasonos A, Barz A, Culkin A, Kris MG, Artz D, et al. Long-term toxicity monitoring via electronic patient-reported outcomes in patients receiving chemotherapy. J Clin Oncol 2007 Dec 01;25(34):5374-5380. [doi: 10.1200/JCO.2007.11.2243] [Medline: 18048818]
14. Cowan R, Suidan R, Andikyan V, Rezk Y, Einstein H, Chang K, et al. Electronic patient-reported outcomes from home in patients recovering from major gynecologic cancer surgery: A prospective study measuring symptoms and health-related quality of life. Gynecol Oncol 2016 Nov;143(2):362-366 [FREE Full text] [doi: 10.1016/j.ygyno.2016.08.335] [Medline: 27637366]
15. Judson TJ, Bennett AV, Rogak LJ, Sit L, Barz A, Kris MG, et al. Feasibility of long-term patient self-reporting of toxicities from home via the Internet during routine chemotherapy. J Clin Oncol 2013 Jul 10;31(20):2580-2585 [FREE Full text] [doi: 10.1200/JCO.2012.47.6804] [Medline: 23733753]
16. Andersen T, Ruland CM. Cancer patients' questions and concerns expressed in an online nurse-delivered mail service: preliminary results. Stud Health Technol Inform 2009;146:149-153. [Medline: 19592825]
17. Børøsund E, Cvancarova M, Ekstedt M, Moore SM, Ruland CM. How user characteristics affect use patterns in web-based illness management support for patients with breast and prostate cancer. J Med Internet Res 2013 Mar 01;15(3):e34 [FREE Full text] [doi: 10.2196/jmir.2285] [Medline: 23454601]
18. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CM. Comparing effects in regular practice of e-communication and Web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. J Med Internet Res 2014 Dec 18;16(12):e295 [FREE Full text] [doi: 10.2196/jmir.3348] [Medline: 25525672]
19. Ruland CM, Jeneson A, Andersen T, Andersen R, Slaughter L, Bente-Schjødt O, et al. Designing tailored Internet support to assist cancer patients in illness management. AMIA Annu Symp Proc 2007 Oct 11:635-639 [FREE Full text] [Medline: 18693913]
20. Ruland CM, Maffei RM, Børøsund E, Krahn A, Andersen T, Grimsbø GH. Evaluation of different features of an eHealth application for personalized illness management support: cancer patients' use and appraisal of usefulness. Int J Med Inform 2013 Jul;82(7):593-603. [doi: 10.1016/j.ijmedinf.2013.02.007] [Medline: 23507561]
21. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
22. 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] [doi: 10.2196/jmir.1639] [Medline: 21212045]
23. Sieverink F, Kelders SM, van Gemert-Pijnen JEWC. Clarifying the Concept of Adherence to eHealth Technology: Systematic Review on When Usage Becomes Adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
24. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9 [FREE Full text] [doi: 10.2196/jmir.1391] [Medline: 21371992]
25. 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] [Medline: 26104044]
26. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The Effectiveness of Prompts to Promote Engagement With Digital Interventions: A Systematic Review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
27. Han JY, Wise M, Kim E, Pingree R, Hawkins RP, Pingree S, et al. Factors Associated with Use of Interactive Cancer Communication System: An Application of the Comprehensive Model of Information Seeking. J Comput Mediat Commun 2010 Apr;15(3):367-388 [FREE Full text] [doi: 10.1111/j.1083-6101.2010.01508.x] [Medline: 21760702]
28. Cushing CC, Steele RG. A meta-analytic review of eHealth interventions for pediatric health promoting and maintaining behaviors. J Pediatr Psychol 2010 Oct;35(9):937-949. [doi: 10.1093/jpepsy/jsq023] [Medline: 20392790]
29. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of Internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014 Jan 16;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]
30. Lu H, Shaw BR, Gustafson DH. Online health consultation: examining uses of an interactive cancer communication tool by low-income women with breast cancer. Int J Med Inform 2011 Jul;80(7):518-528 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.03.011] [Medline: 21530381]
31. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005 Oct 19(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
32. Samoocha D, Bruinvels DJ, Elbers NA, Anema JR, van der Beek AJ. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. J Med Internet Res 2010 Jun 24;12(2):e23 [FREE Full text] [doi: 10.2196/jmir.1286] [Medline: 20581001]

33. 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 14;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
34. Johansen MA, Henriksen E, Horsch A, Schuster T, Berntsen GKR. Electronic symptom reporting between patient and provider for improved health care service quality: a systematic review of randomized controlled trials. part 1: state of the art. J Med Internet Res 2012 Oct 03;14(5):e118 [FREE Full text] [doi: 10.2196/jmir.2214] [Medline: 23032300]
35. Corbett T, Singh K, Payne L, Bradbury K, Foster C, Watson E, et al. Understanding acceptability of and engagement with Web-based interventions aiming to improve quality of life in cancer survivors: A synthesis of current research. Psychooncology 2018 Jan;27(1):22-33. [doi: 10.1002/pon.4566] [Medline: 29044832]
36. Petticrew M. Time to rethink the systematic review catechism? Moving from 'what works' to 'what happens'. Syst Rev 2015 Mar 28;4:36 [FREE Full text] [doi: 10.1186/s13643-015-0027-1] [Medline: 25875303]
37. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Syst Rev 2012 Feb 09;1:2 [FREE Full text] [doi: 10.1186/2046-4053-1-2] [Medline: 22587842]
38. Agency for Healthcare Research and Quality. Systematic Review Data Repository. 2013. URL: https://www.ahrq.gov/cpi/about/otherwebsites/srdr.ahrq.gov/index.html [accessed 2018-11-28] [WebCite Cache ID 74GWIhXTo]
39. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology & Community Health 1998 Jun 01;52(6):377-384. [doi: 10.1136/jech.52.6.377]
40. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M. Guidance on the conduct of narrative synthesis in systematic reviews. ESRC methods programme 2006;15(1):047-071 [FREE Full text] [doi: 10.13140/2.1.1018.4643]
41. McCann L, Maguire R, Miller M, Kearney N. Patients' perceptions and experiences of using a mobile phone-based advanced symptom management system (ASyMS) to monitor and manage chemotherapy related toxicity. Eur J Cancer Care (Engl) 2009 Mar;18(2):156-164. [doi: 10.1111/j.1365-2354.2008.00938.x] [Medline: 19267731]
42. Maguire R, Ream E, Richardson A, Connaghan J, Johnston B, Kotronoulas G, et al. Development of a novel remote patient monitoring system: the advanced symptom management system for radiotherapy to improve the symptom experience of patients with lung cancer receiving radiotherapy. Cancer Nurs 2015;38(2):E37-E47. [doi: 10.1097/NCC.0000000000000150] [Medline: 24836956]
43. Kearney N, Kidd L, Miller M, Sage M, Khorrami J, McGee M, et al. Utilising handheld computers to monitor and support patients receiving chemotherapy: results of a UK-based feasibility study. Support Care Cancer 2006 Jul;14(7):742-752. [doi: 10.1007/s00520-005-0002-9] [Medline: 16525792]
44. Johnston B, Maguire R, Kearney N. Using Mobile Phone Technology to Assess Symptoms In Patients Receiving Palliative Care - The Advanced Symptom Management System (ASyMS (c)-P). In: European Journal of Cancer. 2011 Presented at: 2011 European Multidisciplinary Cancer Congress; Sept. 24, 2011; Stockholm, Sweden.
45. Maguire R, Kearney N, Pedersen V, Ream E, Richardson A. Improving the Symptom Experience of Patients With Lung Cancer Receiving Radiotherapy: Advanced Symptom Management System for Radiotherapy (ASyMS-R). In: European Journal of Cancer. 2011 Presented at: 2011 European Multidisciplinary Cancer Congress; Sept 24, 2011; Stockholm.
46. McCann L, Maguire R, Kearney N, Miller M, Taylor L, Sage M, et al. 8083 ORAL The use of a mobile phone based advanced symptom management system in the home monitoring and symptom management of chemotherapy related toxicities in patients with breast, lung and colorectal cancer: Patients' and clinicans' perceptions for supportive self care. European Journal of Cancer Supplements 2007 Sep;5(4):434. [doi: 10.1016/s1359-6349(07)71585-9]
47. Breen S, Aranda S, Ritchie D, Kofoed S, Dryden T, Maguire R. 'You cannot manage what you cannot measure': Development of a prototype remote monitoring system for haematological cancer patients undergoing chemotherapy. In: Supportive Care in Cancer. 2012 Jun Presented at: International MASCC/ISOO Symposium; June 28-30, 2012; New York p. S69.
48. Cowie J, McCann L, Maguire R, Kearney N, Connaghan J, Paterson C, et al. Real-time management of chemotherapy toxicity using the Advanced Symptom Management System (ASyMS). Journal of Decision Systems 2013 Jan 1;22(1):43-52. [doi: 10.1080/12460125.2012.760269]
49. Steel JL, Geller DA, Kim KH, Butterfield LH, Spring M, Grady J, et al. Web-based collaborative care intervention to manage cancer-related symptoms in the palliative care setting. Cancer 2016 Apr 15;122(8):1270-1282 [FREE Full text] [doi: 10.1002/cncr.29906] [Medline: 26970434]
50. Holzner B, Zabernigg A, Sztankay M, Giesinger J, Kemmler G, Wintner L. Web-based QOL-monitoring for more accurate assessment of symptom burden in cancer patients undergoing chemotherapy. Onkologie 2013 Oct;36:110-111. [Medline: 71279061]
51. Tofthagen C, Kip KE, Passmore D, Loy I, Berry DL. Usability and Acceptability of a Web-Based Program for Chemotherapy-Induced Peripheral Neuropathy. Comput Inform Nurs 2016 Jul;34(7):322-329. [doi: 10.1097/CIN.0000000000000242] [Medline: 27116414]
52. Fishbein JN, Nisotel LE, MacDonald JJ, Amoyal Pensak N, Jacobs JM, Flanagan C, et al. Mobile Application to Promote Adherence to Oral Chemotherapy and Symptom Management: A Protocol for Design and Development. JMIR Res Protoc 2017 Apr 20;6(4):e62 [FREE Full text] [doi: 10.2196/resprot.6198] [Medline: 28428158]

53. Agboola S, Flanagan C, Searl M, Elfiky A, Kvedar J, Jethwani K. Improving outcomes in cancer patients on oral anti-cancer medications using a novel mobile phone-based intervention: study design of a randomized controlled trial. JMIR Res Protoc 2014 Dec 23;3(4):e79 [FREE Full text] [doi: 10.2196/resprot.4041] [Medline: 25537463]
54. Holch P, Gibson A, Holmes M, Rodgers Z, Clayton B, Henry A. Usability testing of an online symptom report and management system in radical prostate radiotherapy (RT) patients: Preliminary findings of the eRAPID RT programme. In: Psycho-Oncology. 2016 Presented at: British Psychosocial Oncology Society annual conference; Mar. 3-4, 2016; Cambridge, UK p. 25.
55. Holch P, Davidson S, Routledge J, Henry A, Franks K, Gilbert A. eRAPID: Electronic self-report and management of adverse-events for radical prostate radiotherapy (RT) patients. In: Radiotherapy and Oncology. 2015 Presented at: ESTRO; April 24-28, 2015; Barcelona, Spain p. 115.
56. Velikova G, Holch P, Bamforth L, Warrington L, Carter R, Absolom K. Development of an integrated online toxicity reporting and management system for oncology: eRAPID (Electronic patient self-Reporting of Adverse-events: Patient Information and aDvice). In: Quality of Life Research. 2014 Oct Presented at: 21st Annual Conference of the International Society for Quality of Life Research; October 15-18, 2014; Berlin, Germany.
57. Holch P, Warrington L, Potrata B, Ziegler L, Absolom K, Hector C. Acceptability of an online system for reporting and managing symptomatic adverse events (eRAPID): Patients views eRAPID is funded by a National Institute for Health Research (NIHR) programme development grant RP-DG-1209-10031; ISCTRN trial number CCT-NAPN-21338. In: Psycho-Oncology. 2013 Presented at: British Psychosocial Oncology Society; January 17-18, 2013; Southampton, UK p. 9-10.
58. Velikova G, Ziegler L, Holch P, Harley C, Keding A, Bamforth L. Development of an electronic platform for patient self-reporting of adverse-events during cancer treatment. In: Asia-Pacific Journal of Clinical Oncology. 2012 Presented at: COSA IPOS Psycho-Oncology 2012; November 13-15, 2012; Brisbane, Australia p. 8.
59. Ziegler L, Harley C, Holch P, Keding A, Bamforth L, Warrington L. Towards safer delivery and monitoring of cancer treatments. Electronic patient self-reporting of adverse-events: Patient information and a advice (eRAPID). In: Psycho-Oncology. 2012 Presented at: British Psychosocial Oncology Society; January 19-20, 2012; Leeds, UK p. 21.
60. Berry DL, Hong F, Halpenny B, Partridge AH, Fann JR, Wolpin S, et al. Electronic self-report assessment for cancer and self-care support: results of a multicenter randomized trial. J Clin Oncol 2014 Jan 20;32(3):199-205 [FREE Full text] [doi: 10.1200/JCO.2013.48.6662] [Medline: 24344222]
61. Berry DL, Blonquist TM, Patel RA, Halpenny B, McReynolds J. Exposure to a patient-centered, Web-based intervention for managing cancer symptom and quality of life issues: impact on symptom distress. J Med Internet Res 2015 Jun 03;17(6):e136 [FREE Full text] [doi: 10.2196/jmir.4190] [Medline: 26041682]
62. Berry DL, Hong F, Halpenny B, Partridge A, Fox E, Fann JR, et al. The electronic self report assessment and intervention for cancer: promoting patient verbal reporting of symptom and quality of life issues in a randomized controlled trial. BMC Cancer 2014 Jul 12;14:513 [FREE Full text] [doi: 10.1186/1471-2407-14-513] [Medline: 25014995]
63. Klasnja P, Hartzler A, Powell C, Pratt W. Supporting cancer patients' unanchored health information management with mobile technology. AMIA Annu Symp Proc 2011;2011:732-741 [FREE Full text] [Medline: 22195130]
64. Klasnja P, Hartzler A, Powell C, Phan G, Pratt W. Health Weaver Mobile: Designing a Mobile Tool for Managing Personal Health Information during Cancer Care. AMIA Annu Symp Proc 2010 Nov 13;2010:392-396 [FREE Full text] [Medline: 21347007]
65. McGee M, Gray P. A handheld chemotherapy symptom management system: results from a preliminary outpatient field trial. Health Informatics J 2016 Jul 25;11(4):243-258. [doi: 10.1177/1460458205055686]
66. Levi F, Arbaud A, Dispersyn G, Simon J, Innominato P, Bossevot R. The European InCASA telecare-telehealth electronic platform (ICT-FP7) for the daily assessment of symptoms, weight, and activity in cancer patients on chronotherapy at home. In: Journal of Clinical Oncology Conference (15 SUPPL. 1). 2012 Presented at: ASCO Annual Meeting; 2012; Chicago, IL.
67. Langius-Eklöf A, Crafoord M, Christiansen M, Fjell M, Sundberg K. Effects of an interactive mHealth innovation for early detection of patient-reported symptom distress with focus on participatory care: protocol for a study based on prospective, randomised, controlled trials in patients with prostate and breast cancer. BMC Cancer 2017 Jul 04;17(1):466 [FREE Full text] [doi: 10.1186/s12885-017-3450-y] [Medline: 28676102]
68. Peltola M, Lehikoinen J, Sippola L, Saarilahti K, Mäkitie AA. A Novel Digital Patient-Reported Outcome Platform for Head and Neck Oncology Patients-A Pilot Study. Clin Med Insights Ear Nose Throat 2016;9:1-6 [FREE Full text] [doi: 10.4137/CMENT.S40219] [Medline: 27721662]
69. McRoy L, Mitra D, Hollis K, Kaye J, Zelnak A, Cheyl J. Abstract OT3-03-01: MADELINE: A prospective observational study of mobile app-based patient reported outcomes in advanced breast cancer. Cancer Res 2017 Feb 14;77(4 Supplement):-. [doi: 10.1158/1538-7445.sabcs16-ot3-03-01]
70. Pusic A, Klassen A, Scott A, Cano S, Shouery M, Basch E. Feasibility and acceptability of patient-reported outcomes data collection for clinical care following breast reconstruction. In: Journal of Clinical Oncology Conference (15 SUPPL. 1). 2012 Presented at: ASCO Annual Meeting; 2012; Chicago, IL.

71. Galligioni E, Piras EM, Galvagni M, Eccher C, Caramatti S, Zanolli D, et al. Integrating mHealth in Oncology: Experience in the Province of Trento. J Med Internet Res 2015 May 13;17(5):e114 [FREE Full text] [doi: 10.2196/jmir.3743] [Medline: 25972226]
72. Passardi A, Rizzo M, Maines F, Tondini C, Zambelli A, Vespignani R, et al. Optimisation and validation of a remote monitoring system (Onco-TreC) for home-based management of oral anticancer therapies: an Italian multicentre feasibility study. BMJ Open 2017 Dec 29;7(5):e014617 [FREE Full text] [doi: 10.1136/bmjopen-2016-014617] [Medline: 28554917]
73. Snyder CF, Blackford AL, Wolff AC, Carducci MA, Herman JM, Wu AW, PatientViewpoint Scientific Advisory Board. Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. Psychooncology 2013 Apr;22(4):895-901 [FREE Full text] [doi: 10.1002/pon.3087] [Medline: 22544513]
74. Della Mea V, De Momi I, Aprile G, Puglisi F, Menis J, Casetta A, et al. Feasibility study of a web application for self-report of anticancer treatment toxicities. Stud Health Technol Inform 2009;150:562-566. [Medline: 19745374]
75. Min YH, Lee JW, Shin Y, Jo M, Sohn G, Lee J, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. J Med Internet Res 2014 May 23;16(5):e135 [FREE Full text] [doi: 10.2196/jmir.3421] [Medline: 24860070]
76. Breen S, Ritchie D, Schofield P, Hsueh Y, Gough K, Santamaria N, et al. The Patient Remote Intervention and Symptom Management System (PRISMS) - a Telehealth- mediated intervention enabling real-time monitoring of chemotherapy side-effects in patients with haematological malignancies: study protocol for a randomised controlled trial. Trials 2015 Oct 19;16:472 [FREE Full text] [doi: 10.1186/s13063-015-0970-0] [Medline: 26481873]
77. Breen S, Aranda S, Ritchie D, Schofield P, Kamateros R, Maguire R. Improving the Management of Chemotherapy Toxicities in Haematological Cancer Patients: A Phase II Randomised Controlled Trial of the Patient Remote Intervention and Symptom Management System (PRISMS). In: Asia-Pacific Journal of Clinical Oncology. 2012 Presented at: ASCO Annual Meeting; 2012; Chicago, IL.
78. Aggarwal S, Topaloglu H. Novel electronic patient reported outcomes tool for prostate cancer patients. In: Value in Health. 2013 May Presented at: ISPOR 18th Annual International Meeting; May 18-22, 2013; New Orleans, LA p. A147.
79. Armstrong KA, Coyte PC, Brown M, Beber B, Semple JL. Effect of Home Monitoring via Mobile App on the Number of In-Person Visits Following Ambulatory Surgery: A Randomized Clinical Trial. JAMA Surg 2017 Jul 01;152(7):622-627 [FREE Full text] [doi: 10.1001/jamasurg.2017.0111] [Medline: 28329223]
80. Armstrong KA, Semple JL, Coyte PC. Replacing ambulatory surgical follow-up visits with mobile app home monitoring: modeling cost-effective scenarios. J Med Internet Res 2014 Sep 22;16(9):e213 [FREE Full text] [doi: 10.2196/jmir.3528] [Medline: 25245774]
81. Rasschaert M, Helsen S, Rolfo C, Van Brussel I, Ravelingien J, Peeters M. Feasibility of an interactive electronic self-report tool for oral cancer therapy in an outpatient setting. Support Care Cancer 2016 Dec;24(8):3567-3571. [doi: 10.1007/s00520-016-3186-2] [Medline: 27025594]
82. Chan M, Ang E, Duong MC, Chow YL. An online Symptom Care and Management System to monitor and support patients receiving chemotherapy: a pilot study. Int J Nurs Pract 2013 Feb;19 Suppl 1:14-18. [doi: 10.1111/ijn.12020] [Medline: 23425375]
83. Chan MF, Ang NKE, Cho AA, Chow YL, Taylor B. Online chemotherapy symptom care and patient management system: an evaluative study. Comput Inform Nurs 2014 Feb;32(2):75-83. [doi: 10.1097/CIN.0000000000000036] [Medline: 24378370]
84. Andikyan V, Rezk Y, Einstein MH, Gualtiere G, Leitao MM, Sonoda Y, et al. A prospective study of the feasibility and acceptability of a Web-based, electronic patient-reported outcome system in assessing patient recovery after major gynecologic cancer surgery. Gynecol Oncol 2012 Nov;127(2):273-277 [FREE Full text] [doi: 10.1016/j.ygyno.2012.07.124] [Medline: 22871467]
85. Head B, Studts J, Bumpous J, Gregg J, Wilson L, Keeney C, et al. Development of a telehealth intervention for head and neck cancer patients. Telemed J E Health 2009 Jan;15(1):44-52 [FREE Full text] [doi: 10.1089/tmj.2008.0061] [Medline: 19199847]
86. Donovan LM, Rueschman M, Weng J, Basu N, Dudley KA, Bakker JP, et al. The effectiveness of an obstructive sleep apnea screening and treatment program in patients with type 2 diabetes. Diabetes Res Clin Pract 2017 Dec;134:145-152. [doi: 10.1016/j.diabres.2017.10.013] [Medline: 29054482]
87. Sun V, Dumitra S, Ruel N, Lee B, Melstrom L, Melstrom K, et al. Wireless Monitoring Program of Patient-Centered Outcomes and Recovery Before and After Major Abdominal Cancer Surgery. JAMA Surg 2017 Sep 01;152(9):852-859 [FREE Full text] [doi: 10.1001/jamasurg.2017.1519] [Medline: 28593266]
88. van den Brink JL, Moorman PW, de Boer MF, Hop WCJ, Pruyn JFA, Verwoerd CDA, et al. Impact on quality of life of a telemedicine system supporting head and neck cancer patients: a controlled trial during the postoperative period at home. J Am Med Inform Assoc 2007;14(2):198-205 [FREE Full text] [doi: 10.1197/jamia.M2199] [Medline: 17213498]
89. van den Brink JL, Moorman PW, de Boer MF, van Bemmel JH, Pruyn JFA, Verwoerd CDA. An information system to support the care for head and neck cancer patients. Support Care Cancer 2003 Jul;11(7):452-459. [doi: 10.1007/s00520-002-0425-5] [Medline: 12707835]

90. van den Brink JL, Moorman PW, de Boer MF, Pruyn JFA, Verwoerd CDA, van Bemmel JH. Involving the patient: a prospective study on use, appreciation and effectiveness of an information system in head and neck cancer care. Int J Med Inform 2005 Oct;74(10):839-849. [doi: 10.1016/j.ijmedinf.2005.03.021] [Medline: 16043392]
91. Newlon C, Hu C, Stratton R, McDaniel A. Design of a Web-Based Symptom Management Intervention for Cancer Patients. 2009 Presented at: Human Centered Design; July 19-24, 2009; San Diego, CA p. 775-784.
92. Sundberg K, Eklöf AL, Blomberg K, Isaksson A, Wengström Y. Feasibility of an interactive ICT-platform for early assessment and management of patient-reported symptoms during radiotherapy for prostate cancer. Eur J Oncol Nurs 2015 Oct;19(5):523-528. [doi: 10.1016/j.ejon.2015.02.013] [Medline: 25813529]
93. Weaver A, Young AM, Rowntree J, Townsend N, Pearson S, Smith J, et al. Application of mobile phone technology for managing chemotherapy-associated side-effects. Ann Oncol 2007 Nov;18(11):1887-1892. [doi: 10.1093/annonc/mdm354] [Medline: 17921245]
94. Prince R, Parente L, Yee A, Powis M, Enright K, Gandhi S. Building “bridges”: Use of participatory design to create an electronic tool to improve management of chemotherapy toxicities. In: Journal of Clinical Oncology Conference. 2016 Presented at: ASCO; June 3-7, 2016; Chicago, IL.
95. Prince RM, Parente L, Soung Yee A, Powis M, Enright K, Gandhi S, et al. Defining user needs for an electronic tool to improve chemotherapy-related toxicity management. JCO 2016 Mar;34(7_suppl):157-157. [doi: 10.1200/jco.2016.34.7_suppl.157]
96. Due J, Christensen H, Vestlev P. Treatment adverse effect registration. In: A PC based electronic patient - hospital communication portal that allows registration of adverse effects of chemotherapy and returns advice and recommandations for action. Cancer Research Conferenceth Annual CTRC AACR San Antonio Breast Cancer Symposium San Antonio, TX United States Conference Start. (9 SUPPL. 1). 2015 Presented at: Cancer Research Conference: 37th Annual CTRC AACR San Antonio Breast Cancer Symposium; 2015; San Antonio, TX p. 37.
97. Nimako K, Lu SK, Ayite B, Priest K, Winkley A, Gunapala R, et al. A pilot study of a novel home telemonitoring system for oncology patients receiving chemotherapy. J Telemed Telecare 2013 Apr;19(3):148-152. [doi: 10.1177/1357633X13483258] [Medline: 23615717]
98. Fann JR, Hong F, Halpenny B, Blonquist TM, Berry DL. Psychosocial outcomes of an electronic self-report assessment and self-care intervention for patients with cancer: a randomized controlled trial. Psychooncology 2017 Nov;26(11):1866-1871 [FREE Full text] [doi: 10.1002/pon.4250] [Medline: 27530529]
99. Bullard E, Coffman E, Bryant A, Bradley J, Tan X, Wood W. Using Electronic Patient Symptom Reporting to Reduce Symptom Burden During Hospitalization for Preparatory Chemotherapy Prior to Hematopoietic Stem Cell Transplant: The Role of the Oncology Nurse. In: Oncology Nursing Forum. 2017 Mar Presented at: Oncology Nursing Society 42nd Annual Congress; May 4-7, 2017; Denver, CO p. 000401160800611.
100. Egbring M, Far E, Roos M, Dietrich M, Brauchbar M, Kullak-Ublick GA, et al. A Mobile App to Stabilize Daily Functional Activity of Breast Cancer Patients in Collaboration With the Physician: A Randomized Controlled Clinical Trial. J Med Internet Res 2016 Dec 06;18(9):e238 [FREE Full text] [doi: 10.2196/jmir.6414] [Medline: 27601354]
101. Falchook A, Tracton G, Stravers L, Fleming M, Reeve B, Basch E. Use of mobile device technology to collect patient-reported symptoms during radiotherapy for head and neck cancer: A prospective feasibility study. In: Journal of Clinical Oncology Conference (15 SUPPL. 1). 2015 Presented at: ASCO; May 29-June 2, 2015; Chicago, IL.
102. Yang M, Thea J, An Y, Yu JB. Digital health application for real-time patient-reported outcomes during prostate radiotherapy. JCO 2016 Jan 10;34(2_suppl):157-157. [doi: 10.1200/jco.2016.34.2_suppl.157]
103. McCorkle R, Ercolano E, Lazenby M, Schulman-Green D, Schilling LS, Lorig K, et al. Self-management: Enabling and empowering patients living with cancer as a chronic illness. CA Cancer J Clin 2011;61(1):50-62 [FREE Full text] [doi: 10.3322/caac.20093] [Medline: 21205833]
104. Kofoed S, Breen S, Gough K, Aranda S. Benefits of remote real-time side-effect monitoring systems for patients receiving cancer treatment. Oncol Rev 2012 Mar 05;6(1):e7 [FREE Full text] [doi: 10.4081/oncol.2012.e7] [Medline: 25992209]
105. McAlpine H, Joubert L, Martin-Sanchez F, Merolli M, Drummond K. A systematic review of types and efficacy of online interventions for cancer patients. Patient Educ Couns 2015 Mar;98(3):283-295. [doi: 10.1016/j.pec.2014.11.002] [Medline: 25535016]
106. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
107. Solomon M, Wagner SL, Goes J. Effects of a Web-based intervention for adults with chronic conditions on patient activation: online randomized controlled trial. J Med Internet Res 2012 Feb 21;14(1):e32 [FREE Full text] [doi: 10.2196/jmir.1924] [Medline: 22353433]
108. Hibbard J, Mahoney E, Sonet E. Does patient activation level affect the cancer patient journey? Patient Educ Couns 2017 Jul;100(7):1276-1279. [doi: 10.1016/j.pec.2017.03.019] [Medline: 28330715]
109. Papadopoulou C, Kotronoulas G, Schneider A, Miller MI, McBride J, Polly Z, et al. Patient-Reported Self-Efficacy, Anxiety, and Health-Related Quality of Life During Chemotherapy: Results From a Longitudinal Study. Oncol Nurs Forum 2017 Dec 01;44(1):127-136. [doi: 10.1188/17.ONF.127-136] [Medline: 27991612]

110. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007 Aug;42(4):1443-1463 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00669.x] [Medline: 17610432]
111. Marshall R, Beach MC, Saha S, Mori T, Loveless MO, Hibbard JH, et al. Patient activation and improved outcomes in HIV-infected patients. J Gen Intern Med 2013 May;28(5):668-674 [FREE Full text] [doi: 10.1007/s11606-012-2307-y] [Medline: 23288378]
112. Rask KJ, Ziemer DC, Kohler SA, Hawley JN, Arinde FJ, Barnes CS. Patient activation is associated with healthy behaviors and ease in managing diabetes in an indigent population. Diabetes Educ 2009;35(4):622-630. [doi: 10.1177/0145721709335004] [Medline: 19419972]
113. Begum N, Donald M, Ozolins IZ, Dower J. Hospital admissions, emergency department utilisation and patient activation for self-management among people with diabetes. Diabetes Res Clin Pract 2011 Aug;93(2):260-267. [doi: 10.1016/j.diabres.2011.05.031] [Medline: 21684030]
114. Beller EM, Chen JK, Wang UL, Glasziou PP. Are systematic reviews up-to-date at the time of publication? Syst Rev 2013 May 28;2:36 [FREE Full text] [doi: 10.1186/2046-4053-2-36] [Medline: 23714302]
115. Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice. CA Cancer J Clin 2012;62(5):337-347 [FREE Full text] [doi: 10.3322/caac.21150] [Medline: 22811342]
116. Boger E, Ellis J, Latter S, Foster C, Kennedy A, Jones F, et al. Self-Management and Self-Management Support Outcomes: A Systematic Review and Mixed Research Synthesis of Stakeholder Views. PLoS One 2015;10(7):e0130990 [FREE Full text] [doi: 10.1371/journal.pone.0130990] [Medline: 26162086]

Abbreviations

ASyMs: Advanced Symptom Management System
CASSY: Comprehensive Electronic Cancer Support System for the Treatment of Cancer Related Symptoms
CTCAE: Common Toxicity Criteria Adverse Events
ESRA-C: Electronic Self-Report Assessment-Cancer
HCP: health care professional
PICOS: Population, Intervention, Comparator, Outcomes, Study design
PROM: patient-reported outcome measure
PROSPERO: International Prospective Register of Systematic Reviews
QAS: Quality Assessment Score
QoL: quality of life
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 25.04.18; peer-reviewed by L Hess, M Krzyzanowska, V Kukreti, R Collado-Borrell; comments to author 25.08.18; revised version received 20.09.18; accepted 10.10.18; published 24.01.19.

Please cite as:
Warrington L, Absolom K, Conner M, Kellar I, Clayton B, Ayres M, Velikova G
Electronic Systems for Patients to Report and Manage Side Effects of Cancer Treatment: Systematic Review
J Med Internet Res 2019;21(1):e10875
URL: http://www.jmir.org/2019/1/e10875/
doi:10.2196/10875
PMID:30679145

Review

Consumer Health Information Technology in the Prevention of Substance Abuse: Scoping Review

Apoorva Milind Pradhan[1*], BAMS, MPH; Leah Park[1*], MS, PharmD; Fadia T Shaya[1*], MPH, PhD; Joseph Finkelstein[2*], MD, PhD

[1]Department of Health Services Research, University of Maryland School of Pharmacy, Baltimore, MD, United States

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

[*]all authors contributed equally

Corresponding Author:
Fadia T Shaya, MPH, PhD
Department of Health Services Research
University of Maryland School of Pharmacy
220 Arch Street, 12th Floor, 01-204
Baltimore, MD, 21201
United States
Phone: 1 410 706 5392
Email: fshaya@rx.umaryland.edu

Abstract

Background: Addiction is one of the most rapidly growing epidemics that currently plagues nations around the world. In the United States, it has cost the government more than US $700 billion a year in terms of health care and other associated costs and is also associated with serious social, physical, and mental consequences. Increasing efforts have been made to tackle this issue at different levels, from primary prevention to rehabilitation across the globe. With the use of digital technology rapidly increasing, an effort to leverage the consumer health information technologies (CHITs) to combat the rising substance abuse epidemic has been underway. CHITs are identified as patient-focused technological platforms aimed to improve patient engagement in health care and aid them in navigating the complex health care system.

Objective: This review aimed to provide a holistic and overarching view of the breadth of research on primary prevention of substance abuse using CHIT conducted over nearly past five decades. It also aimed to map out the changing landscape of CHIT over this period.

Methods: We conducted a scoping review using the Arksey and O'Malley's modified methodological framework. We searched 4 electronic databases (PubMed, Cochrane, Scopus, and EMBASE). Papers were included if the studies addressed the use of CHIT for primary prevention of substance abuse and were published in English between 1809 and 2018. Studies that did not focus solely on primary prevention or assessed additional comorbid conditions were eliminated.

Results: Forty-two papers that met our inclusion criteria were included in the review. These studies were published between 1970 and 2018 and were not restricted by geography, age, race, or sex. The review mapped studies using the most commonly used CHIT platforms for substance abuse prevention from mass media in the 1970s to mobile and social media in 2018. Moreover, 191 studies that were exclusively focused on alcohol prevention were excluded and will be addressed in a separate paper. The studies included had diverse research designs although the majority were randomized controlled trials (RCT) or review papers. Many of the RCTs used interventions based on different behavioral theories such as family interactions, social cognitive theories, and harm-minimization framework.

Conclusions: This review found CHIT platforms to be efficacious and cost-effective in the real-world settings. We also observed a gradual shift in the types and use of CHIT platforms over the past few decades and mapped out their progression. In addition, the review detected a shift in consumer preferences and behaviors from face-to-face interactions to technology-based platforms. However, the studies included in this review only focused on the aspect of primary prevention. Future reviews could assess the effectiveness of platforms for secondary prevention and for prevention of substance abuse among comorbid populations.

(J Med Internet Res 2019;21(1):e11297) doi:10.2196/11297

KEYWORDS

consumer health information technology; primary prevention; substance abuse; review

Introduction

Background

Addiction has been identified as the most neglected disease in the United States, with nearly 40 million Americans over the age of 12 years meeting the clinical criteria for addiction involving nicotine, alcohol, or other drugs [1]. In addition, it is also estimated that nearly 80 million people in the country are *risky substance users*, meaning that although they are not addicted, they use tobacco, alcohol, and other drugs in ways that could be harmful and threaten public health and safety [1]. Thus, addiction has been established as a growing epidemic. In 2017, the American government spent in excess of US $740 billion in costs related to crime, lost work, and health care, and costs are growing exponentially [2-6].

However, the burden of drug and alcohol use and abuse is not just limited to the United States. The World Health Organization (WHO) estimates the global burden of disease related to drug and alcohol abuse to be nearly 5.4% [3]. Drug use is associated with grave long-term and short-term health implications and is recognized as one of the avoidable causes of mortality [7]. According to the National Institute on Drug Abuse, in the United States, nearly 64,070 people died of drug overdose in 2016, and this number is rapidly increasing [8]. Given the dire medical, social, and economic consequences associated with increasing drug use, the governments and institutions around the world have been working tirelessly to develop strategies to combat this drug epidemic.

Among the many interventions developed to address the drug epidemic, prevention strategies have shown to be effective to reduce the onset of drug use and risk of drug abuse [9]. Prevention in the field of addiction and substance abuse may be classified as primary, secondary, and tertiary [10]. Primary prevention could be defined as efforts to improve health and forestall the onset of substance abuse and delay the first use (WHO). Primary prevention can be targeted toward the general population, individuals who are at risk of substance abuse, or those who have signs indicating predisposition to developing addiction [11]. As substance abuse can lead to other medical and social problems and have a negative impact not only on individuals, their families, and the society (Substance Abuse and Mental Health Services Administration), effective primary prevention strategies can have beneficial cascading effects in the long run [12].

Among the various strategies implemented for prevention, consumer health information technology (CHIT) has emerged as a potentially effective way to prevent and treat substance abuse [13]. CHIT has been defined in several ways [14-18]. For the purpose of our review, we decided to adopt a combination of 2 definitions of CHIT by Or et al and Tao et al to broaden our scope of work [14,17,18]. Or et al have defined CHIT as "patient-focused interactive web- or technology-mediated applications that are designed to improve information access and exchange, enhance decision making, provide social and emotional support, and facilitate behavior changes that promote health and wellbeing" [14,18], whereas Tao et al have defined it as "consumer-centered electronic tools, technologies, apps, or systems that are interacted with directly by health consumers to provide them with data, information, recommendations, or services for promotion of health and health care" [17]. CHIT has experienced an exponential growth during the recent decades [13]. The traditional CHIT platforms for health promotion such as educational audio and video materials, along with the rapid proliferation of new modalities, take advantage of the wide accessibility of health-related content via internet, mobile phones, and social media [19]. By 2016, nearly 88% Americans were using the internet [20], 92% a cell phone, and 76% social media [21]. Given the reach of these modalities, technologies such as the internet and mobile phones are viewed as a promising platform for affecting assessment, prevention, treatment, and recovery of substance abuse disorders by national and global organizations [22].

In the past few decades, multiple research studies have been conducted on the use of CHIT in the prevention of substance abuse, but the reflective step of looking broadly across this vast research corpus is yet to be undertaken. Previous reviews on CHIT and substance abuse have assessed the impact of CHIT with narrow focuses in terms of targeted study populations and specific types of technology [23-25]. Moreover, 1 systematic review evaluated benefits, potentials, and shortcomings of recent technology such as social media and mobile apps as an intervention for substance use among those who also have HIV. This study concluded the new technology is well accepted and has good feasibility with great potential for educating people on sensitive topics [25]. Another study investigated the benefits of technology on prevention and treatment of substance use among young people and reported that technology is particularly effective in both prevention and treatment regardless of the stage of substance abuse [23]. Moore et al reported that the use of computers in the prevention of drug abuse was effective in reducing substance use and improving knowledge, leading to greater motivation to change behavior and was well accepted by users [24].

Although these reviews presented the impact of CHIT on substance abuse from various perspectives, there has not been a comprehensive review of how interventions using CHIT have shaped primary prevention of substance abuse over the years. Furthermore, some of the previous reviews were limited to studies that used technologies used in recent decades [23,25]. Before the advent of mobile phones and internet, other types of technologies, including phone, television, radios, and videos, played a prominent role in prevention of substance abuse [26-30]. Although some of these technologies from the previous decades are no longer in widespread use, there are valuable lessons that can be learned from which past strategies using technology-based interventions were effective or ineffective and how technology has evolved over the years. Due to the lack of studies that provided overall assessment of effectiveness of CHIT on primary prevention of substance abuse, we focused

the scope of our review on primary prevention of various types of substance abuse and identified studies that provided prevention efforts to reduce the new onset of substance abuse.

Objective

In this study, we present a scoping review of the breadth of research over the past few decades, specifically with the use of CHIT in the primary prevention of substance abuse. The objective of this review was to describe the use of CHIT in the primary prevention of substance abuse over the last five decades and examine the changes and developments in the types of CHIT employed for this effort. Our goal was also to summarize these preventive approaches and report lessons learned from these studies.

Methods

After considering the multiple systematic approaches that are used for the review of published literature, we decided to undertake a scoping review to map out the changing trends in the use of CHIT in the substance abuse prevention landscape over the past few decades. The scoping review methodology is more commonly known as mapping, a process of summarizing the range of evidence to convey the depth and the breadth of the published literature in a particular field of interest. Unlike systematic reviews and meta-analysis, scoping reviews are neither limited by the type of study under consideration nor do they evaluate their quality. Yet, it enables the researcher to examine the extent, range, and nature of research activity; determine whether a full systematic review would be of value; summarize and disseminate the research findings; and identify gaps in the literature [31-34].

In designing our scoping review, we used Arksey and O'Malley's pioneering framework and incorporated recent scoping review publications as well. Arksey and O'Malley's scoping review framework outlines a 5-stage approach, which was further adapted and modified to some extent by others to develop a more feasible approach for reviewing such a vast body of literature [31,32]. The 5 steps are each discussed below.

Identifying the Research Questions

The growing drug abuse epidemic in the United States underscores the need for exploring new approaches to prevention. The ubiquitous nature of CHIT in our day-to-day lives presents an opportunity to study its potential as a tool for substance abuse prevention. Our intent, thus, was to learn the extent of the present use of CHIT platforms in substance abuse prevention; however, the scope of this review was only limited to primary prevention, and not secondary prevention. In addition, we intended to explore the best methods to leverage CHIT platforms in the future among high-risk individuals for primary prevention. Our goal was to examine the following: (1) extent, range, and nature of the evidence; (2) identify gaps in the literature; and (3) summarize and disseminate this information to guide practice and policy. Following Levac et al's suggestion to enhance and advance Arksey and O'Malley framework, our team clarified and linked the purpose and the research question from the beginning of this study. To avoid leading with a *highly focused research question*, we asked a sufficiently broad question: what is the role and scope of CHIT in the primary prevention of substance abuse? [31,32]. To further guide our review, we formulated 4 subquestions:

1. What are the demographics related to substance abuse disorder studied?
   - Locations—study site—the United States, the United Kingdom, and multicountry
   - Demographics of the study population and sample size of the study
   - Length of observation—long-term impact versus short-term impact
2. What is the type of intervention and the behavioral framework, if any, used?
   - Intervention—primary prevention
   - Behavioral framework—transtheoretical model, motivational interviewing, brief intervention, acceptance and commitment therapy, and psychoeducation
3. What is the type of CHIT used?
   - CHIT—desktop, tablet, mobile phone, internet, interactive voice response, video or movies, video recording or audio recording, and radio
   - Social media—Baidu Tieba, Facebook (and its associated Facebook Messenger), Gab, Google+, Myspace, Instagram, LinkedIn, Pinterest, Tumblr, Twitter, Viber, VK, WeChat, Weibo, WhatsApp, Wikia, Snapchat, and YouTube
4. What are the major takeaways from the literature in terms of outcomes and is the intervention effectiveness evident within the literature?
   - Outcomes—change in knowledge, attitudes, and behaviors
   - Effectiveness—intervention outcomes presented by the author(s) and their suggestions for future research

Identifying Relevant Studies

Arksey and O'Malley in their study had emphasized the need to be comprehensive while conducting a scoping review [31]. At this stage, our team deliberated and decided on the various search terms, databases, search strategies, and eligibility criteria. To maintain a broad approach, we did not limit the inclusion of studies by the type of substance used. With the aid of a librarian, we searched electronic databases such as PubMed, Scopus, EMBASE, and the Cochrane library. Various search terms and their combinations were used to identify relevant studies, which discussed the use of CHIT in substance abuse prevention: "technology, internet, cell phone, multimedia, computer-assisted, telemedicine, social media, internet, web-based, etc. with prevention and control, preventive health services and substance-related disorders, substance abuse, substance misuse, drug addiction." An exhaustive list of the search terms used can be found in the Multimedia Appendix 1.

Eligibility Criteria

The following inclusion criteria were used to guide the search and were also used for reviewing papers:

- Publication language—English
- Only limited to human subjects
- Time range—from 1809 to January 2018
- All age groups
- Review papers included—research studies, systematic reviews, meta-analysis, narrative reviews, observational studies, randomized control trials (RCT), qualitative studies, completed clinical trials, and dissertations and working papers
- Reviews, including but not limited to developed countries, given the growing drug abuse presence all around the world. Studies were included from the United States, the United Kingdom, Canada, Europe, Middle East, South America, Southeast Asia, New Zealand, and Australia
- Studies that address the role of CHIT in primary prevention of only substance abuse—defined as strategies to prevent initiation of substance abuse

Exclusion criteria are as follows:

- Journal papers that are not research studies or reviews (ie, those besides the ones defined in the inclusion list) such as editorial reviews, commentaries, opinion papers, and book reviews
- Research targeting secondary prevention strategies such as treatment, maintenance, relapse, and interventions
- Research studies that lacked the use of CHIT as a part of their interventions for primary prevention
- Research studies aimed at prevention of substance abuse among individuals with comorbidities such as HIV and risky health behaviors
- Research studies conducted in special populations such as cancer patients and patients with HIV, AIDS, or other sexually transmitted diseases (STDs)

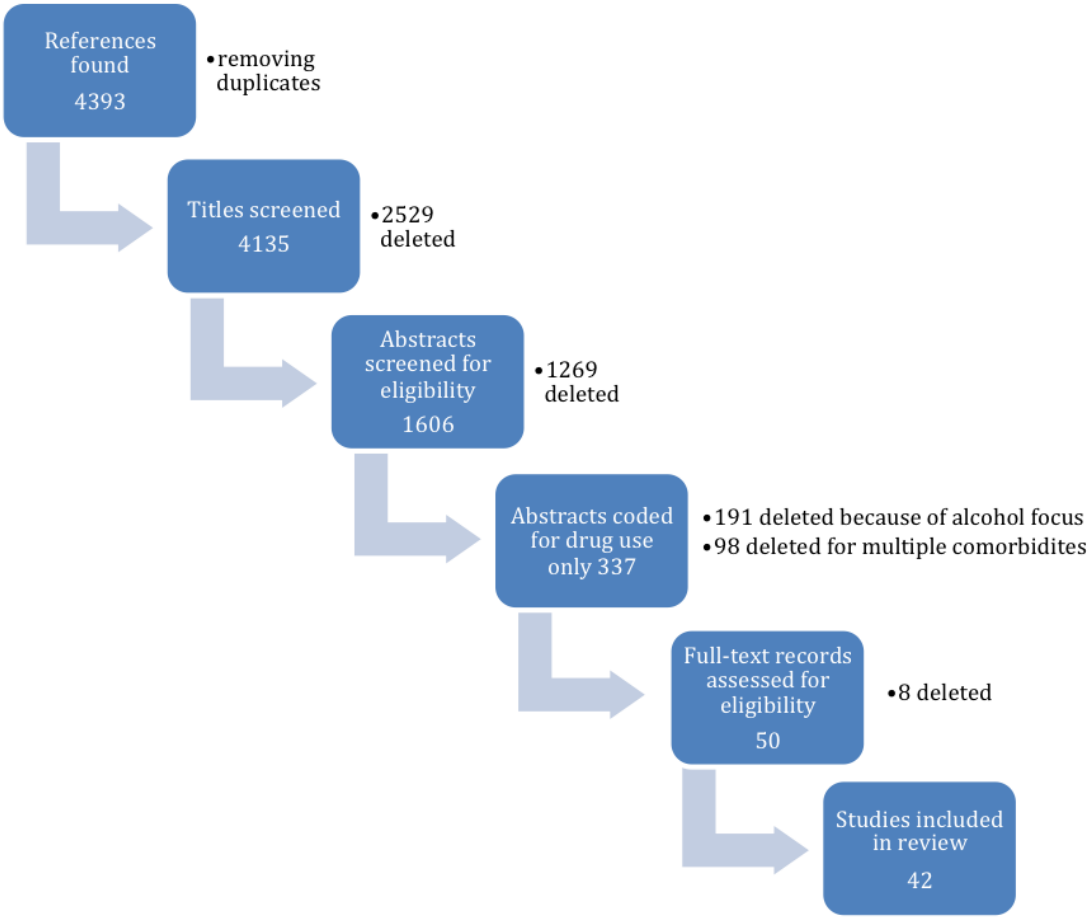
Results

Study Selection

The literature search yielded 4393 papers. Following the search, the study selection was conducted in 2 parts. First, a single reviewer conducted a title screening process based on the inclusion and exclusion criteria. At this stage of screening, any ambiguity after reviewing the title, with regard to the context of the study, did not eliminate it from being considered for the next step. After completion of this step, 1606 papers were identified; then based on the eligibility criteria, 2 reviewers independently conducted abstract coding. On completion of abstract coding, the inclusion of the paper for full-text review was determined by corroboration from both the reviewers. Given the objective of this review, we wanted to assess the overall breadth of the literature on alcohol; thus, our initial yield for all studies included alcohol whether with or without other substances. On the initial screen, we only kept those studies that considered alcohol along with other substances to make sure that we address the impact of CHIT in prevention of all types of substance abuse without our conclusions being skewed because of any single substance of abuse. We think that it is important to note that about 79.2% (191/241) of all the overall CHIT literature on substance use addressed only alcohol, whereas 20.7% (50/241) addressed other substances with or without alcohol. Our intent was to include studies addressing alcohol to the extent that they also addressed the concurrent use of other substance. Given the epidemiology and suggestive evidence of alcohol being a gateway to other drugs, we decided to separately synthesize the studies focused singularly on alcohol in an independent review altogether. Hence, for the purpose of this review, papers focusing solely on alcohol abuse prevention were excluded. We only included studies that focused solely on substance abuse prevention. Studies that looked at substance abuse prevention in conjunction with other morbidities such as prevention of HIV, or other areas of education such as undertaking risky sexual behaviors, were excluded from this review.

Following abstract coding, 50 papers were included for a full-text review. At the end of the full-text review, 42 papers were found to meet all our inclusion criteria and were included in the study. The specific steps of study selection and the number of papers included and excluded in each step can be seen in Figure 1 (flowchart for literature search and inclusion in Multimedia Appendix 2). After the study selection, information relevant to each of the main review questions was extracted and analyzed. We developed a standardized table (Multimedia Appendix 2) using these 42 papers selected to assess the different forms of CHIT used and their impact on the prevention of substance use. We also used this table to identify the different underlying behavioral frameworks most commonly at play in these interventions. Additionally, we also created a table to compile all the papers included in the reviews that are enlisted in this scoping review (Multimedia Appendix 3 [35-205]).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart for literature search process and inclusion.



Locations

Most studies included in the review were conducted in the United States. Countries other than the United States were Canada [206-208], Australia [207-212], Brazil [213], the Netherlands [214], New Zealand [29], Norway [209,215,216], the United Kingdom [28,209,217], Germany [207], and Switzerland [218]. Moreover, 1 systematic review included 7 studies from the United States, and 1 from Norway [216]. Another study reviewed trials that took place in Australia, the Netherlands, the United Kingdom, and the United States [209]. In addition, 1 RCT took place in the United States and Canada [206]. A total of 6 studies included in the review were conducted in Australia. Within the United States, 1 randomized study recruited participants in West Virginia and Ohio [219]. Another RCT in the United States chose participants from 19 states and included Asian populations [220]. Furthermore, 1 study using Monitoring the Future survey data included nationally representative sample of students from 48 states [221]. Studies from other states in the United States recruited participants from California [30,222], New York [223], New Jersey [223], Connecticut [223], Texas [224], Kansas [222], Michigan [225], Missouri [222], and South Carolina [27]. Moreover, 1 study reported having participants mostly from rural communities in South Carolina [27]. Another study chose participants from schools in a semirural community in Michigan [225].

Participants

We did not exclude any studies in the review based on the demographic characteristics of participants. Therefore, study participants varied in ages, racial and ethnic backgrounds, and socioeconomic status. A total of 13 studies assessed school- or college-based intervention programs, and thus, participants were students, students with parents, or teachers [26,30,210,213,217,221,222,224-229]. Moreover, 1 quasi-experimental study included participants enrolled in vocational schools [218]. Similarly, 30 studies included only young people, adolescents, or school-aged children from 11 to 24 years of age. However, some systematic reviews assessed studies across all ages, both children and adults [207,230-233]. Some studies included only racial and ethnic minorities such as Asians [220], Hispanics [223], and African Americans [223]. Although some studies did not specify racial background of participants, 2 studies reported having study subjects from various races, including white, African American, Hispanic, and Asian [30,224]. Moreover, 1 study in the United Kingdom having subjects from 7 schools included 1 school with predominantly black students and the rest with mostly white students [217]. Another UK school-based study reported their participants were chosen to have a balance between both sexes, residence in rural and urban areas, and varying intellectual abilities [28]. A total of 4 studies specifically evaluated interventions on adolescent girls [220,223,234,235]. Of the 4 studies, 3 studies used a mother-daughter prevention approach [220,223,227], assessing the impact of programs both on

adolescent girls and their mothers. Moreover, 1 study selected participants only from economically disadvantaged African American adolescents [236].

Sample Size

A range of study sample size was included in the review. Overall, 1 systematic review had a total of 52,746 individuals as participants from 8 studies [215]. Other reviews included study sample sizes as small as 38 [207] to as large as 8352 [209]. Mother and daughter programs ranged from sample sizes of 206 to 916 girls and their mothers [220,223,234]. The nationally representative survey data included 337,918 cases [221]. Quasi-experimental studies also included both small (n=26) and large (n=2882) sample sizes [30,236]. The RCTs reported larger sample sizes in general, ranging from 179 to 2332 participants [229].

Comparators

Of the 42 studies included in this review, Table 1 shows the different CHIT platforms used by the studies included. Moreover, 4 studies (Schuman et al, 1971; Milne et al, 1975; Pickens, 1984; and Eiser et al, 1988) looked at the effects of films as a part of educational programs to promote discussion around the areas of substance use [28,29,217,225]. The Schuman et al's (1971) study was a cross-sectional study conducted to evaluate the results of a field test implemented to emphasize on aspects such as motivations governing drug behaviors as opposed to drug facts [225]. The Milne et al's (1975) study used a pre-post study design [28], whereas the Eiser et al's (1988) study used an RCT design to assess the effectiveness of films as a medium of drug education [217]. The Pickens' (1984) study was a literature review aimed at assessing the effectiveness of films in drug education compared with other forms of media [29].

A total of 10 studies looked at the use of different mass media interventions in general. Of these, 7 studies (Wallack, 1980; Wallack, 1981; Bandy et al, 1983; Flay et al, 1983; Flay et al, 1986; Brinn et al, 2010; and Carson et al, 2017) were literature reviews, which evaluated the use and effectiveness of mass media as a tool for substance abuse prevention [215,216,226,231,237-239]. The Barcus et al (1975) and the Kinder's (1975) studies examined the impact of mass media on attitudes associated with increased substance use [230,240]. Although majority of the studies evaluated the impact of mass media on multiple drug use, the Brinn et al (2010) and the Carson et al's (2017) studies looked specifically into its role in smoking prevention [215,216]. The Miller et al's (1981) study used a cross-sectional survey design to compare effectiveness of media platforms such as television, radio, and newspaper in dissemination of substance abuse education [27]. Moreover, 5 studies looked specifically into the role of television and radio as modes of interventions. In addition, 3 studies (Feingold et al, 1977; Sussman et al, 1987; and Brannon et al, 1989) used a quasi-experimental design [26,30,222]. The Sussman's study and the Brannon's study used a school-based television program format [30,222], whereas Feingold (1977) and Terry-McElrath et al (2011) used television advertisements as their mode of intervention [26,221]. The Johnson et al's (1989) study was the only one that reviewed the strategies and research efforts in the use of radio and television [232].

Post 2009, there was a notable surge in the number of computer- and internet-based interventions and a subsequent decrease in the number of mass media–based interventions. There were in total 18 studies that used Web-based and internet-based programs for interventions, whereas 5 studies used the desktop-enabled software or CD-ROM-based programs for interventions. Moreover, 13 of these studies used an RCT design and aimed to evaluate the effectiveness of computer-based substance abuse prevention programs. The Hansen et al, Newton et al (2014 and 2016), Christoff Ade et al (2015), and Andrews et al (2011 and 2014) studies evaluated the comparative effectiveness of computer-delivered prevention or screening programs with those termed as usual or traditional models of delivery [210-213,228,229,241]. In addition, 4 studies performed by Schinke et al compared the effectiveness of a mother-daughter–based program with a control group involving no intervention [220,223,227,234]. Furthermore, 2 studies by Schwinn et al tested the effect of an internet-based gender-specific drug prevention program between girls with and without an intervention [206,235]. Another 2 studies used the quasi-experimental pre-post study design, the Klisch et al's (2013) study compared the effectiveness between 2 different Web-based interactive programs [224], and the Moncher et al (1989) assessed the efficacy of the computer-delivered prevention programs [236]. Of the remaining studies, 5 studies were systematic reviews aimed at expanding the base of research and synthesizing the effectiveness of computer- and Web-based prevention programs.

The review by Carson et al (2017) and the randomized trial by Schwinn et al (2017) specifically included the effectiveness of social media as a component of the prevention programs [216,235]. In addition, a quasi-experimental (pre-post assessment) study by Haug et al (2017) and 2 systematic reviews by Jiang et al (2017) and Kazemi et al (2017) evaluated and critiqued the effectiveness of mobile phone–based prevention programs [218,233,242].

Table 1. Technology used as intervention in the 42 studies reviewed

Technology	Studies, n (%)
Computer	6 (10)
CD ROM	5 (9)
Film	4 (6)
Internet	18 (31)
Mass media	6 (10)
Mobile	3 (5)
Radio	6 (10)
Television	11 (19)

Length of Observation

Of the 42 studies reviewed here, 17 studies were in the form of systematic reviews spanning over the past 60 years. Of the remaining 27 studies, 8 studies had a short-term follow-up period of 1 month or less (Schuman et al; Milne et al, 1975; Feingold et al, 1977; Eiser et al, 1988; Moncher et al, 1989; Andrews et al, 2011; Deitz et al, 2011; and Klisch et al, 2013) [26,28,217,219,224,225,236,241]. Most of these studies used a cross-sectional pre-post assessment format. The long-term follow-up periods for most of the remaining studies ranged between 6 months (Schwinn et al, 2010; Champion et al, 2016; and Haug et al, 2017) to 1 year (Sussman et al, 1987; Schinke et al, 2009; Fang et al, 2010; Newton et al, 2014; Newton et al, 2016; and Schwinn et al, 2017) [30,206,210,212,218,220,234,235]. In addition, 2 studies, Schinke et al's (2009) that evaluated computer-delivered program in preventing abuse among adolescent girls and Andrews et al's (2014) that assessed the long-term efficacy of a tobacco prevention program, had a follow-up period of 2 years [227,229]. The study by Christoff Ade et al in 2015 compared the efficacy of 3 different interventions, including a computer-delivered one, and followed its participants for 3 months [213]. The longest study period in this review was 10 years (Terry-McElrath et al, 2011); it evaluated the impact of antidrug advertisement exposure and campaign-specific exposure on the attitudes, beliefs, and behaviors among youths from 1995 to 2006 [221].

Outcomes and Results

Of the 4 studies that looked at films as the mode of intervention delivery, the Schuman et al's (1971) study found no significant difference in the identification of drug clues by geographical or socioeconomic differences [225]. It also found a large gap in perceptions about drugs among faculty and students [225]. The Milne et al's (1975) study found no significant difference in knowledge and attitudes toward drug use. Instead, results showed that students who believed drug use had social advantages also held onto the concept that the dangers of drug abuse were over exaggerated, a finding that thereby emphasizes the need for drug education [28]. The review study conducted by Pickens et al (1984) did not find film interventions superior to nonfilm approaches and found that the short-term impact of film interventions did not last in long-term follow-up studies [29]. However, the Eiser et al's (1988) study showed that an entertaining drug prevention film might be more effective in leading students to reject dangerous substances. In contrast, the students who viewed the educational film regarded both illegal and legal drugs to be similarly dangerous and addictive [217].

Studies that used mass media, radio, and television as modes of intervention found that neither of these platforms as stand-alone were adequate to bring about a change in the overall attitudes and behaviors of people who engage in substance use. The Barcus et al (1975), Wallack et al (1980), Wallack et al (1981), and Flay et al's (1983) studies found that mass media alone is not sufficient to affect behavioral changes and that it needs to be supplemented by school- or community-based prevention programs [215,216,230,231,237-240]. In addition, the literature reviews conducted on the use of mass media by Kinder (1975), Bandy et al (1983), Brinn et al (2010), Carson et al (2017), and a study by Sussman et al (1987) found either inconclusive or conflicting end results pertaining to the use of mass media in disseminating drug-related information and bringing about attitude changes [30,215,216,230,231]. Some studies (eg, Feingold et al, 1977) also found a boomerang effect of the use of television and radio, and antidrug messages were found to potentially lead to drug use [26]. Another exploratory study by Miller et al (1981) evaluated the comparative effectiveness among different mass media platforms such as television, radio, and newspaper and found that it depended on the demographics of the target audience: the results varied by race, sex, and geographical area [27]. Only the Brannon et al's study (1989) specifically looked at the effectiveness of television as a delivery format found it to have higher classroom participation rates, greater satisfaction, and higher perceived effectiveness for a combined television and classroom program, thus concluding it to be a viable option for wider implementation [222].

Post 2009, there was a notable increase in the number of studies that used computer- and Web-based interventions for substance use prevention. A total of 4 studies conducted by Schinke et al between 2009 and 2011 used mother-daughter dyads from different races to study the effectiveness of computer-delivered interventions based on the family interaction theory [220,223,227,234]. All studies found significant reductions in risk factors, drug use, and an increase in the protective factors. Some studies also showed improvements in the quality of mother-daughter relationships. Moreover, 2 studies conducted by Schwinn et al in 2010 and 2017 used gender-specific interventions for girls using internet and social media platforms.

These studies found reduced 30-day alcohol, marijuana, poly-drug, and total substance use at 6-month and 1-year follow-up [206,235]. The 2017's study also found material changes in the cognition and skills that are empirically linked to drug use risks [235]. In addition, 2 studies by Andrew et al in 2011 and 2014 analyzed the short- and long-term efficacy of *Click City tobacco intervention* and found that the intervention had the potential to significantly postpone or prevent the initiation of cigarette use and regular smoking among students. In addition, although in the short term, the intervention showed moderate effectiveness at changing intentions to use smokeless tobacco in the future, the effect did not persist in the long term [229,241]. Another study (Deitz et al, 2011) that evaluated the effect of the Smart Rx Web program found that it significantly increased participants' knowledge of proper prescription drug use and improved their self-efficacy in ability to manage and adhere to appropriate treatments [219].

Multiple studies identified in this review had used school-based programs for the delivery of Web-based interventions; 3 of these conducted by Newton et al between 2013 and 2016 in Australia used the climate schools format for the prevention of use of drugs such as cannabis, alcohol, ecstasy, and new psychoactive substances (NPS) [210-212]. These studies not only found evidence that internet-based preventive interventions significantly decreased substance use but also demonstrated that they could concurrently reduce associated risk factors in adolescents. However, the intervention neither significantly changed binge drinking and cannabis nor ecstasy and NPS use in the short term; the effects of these interventions were only apparent after 12 months, thereby showing a time-delayed effect, which could be attributed to the time required by the students to experience and implement the strategies learned [212]. The Hansen et al's study in 2009 that evaluated the efficacy of Web-based components to facilitate program implementation concluded that school-based prevention programs could benefit from adding Web-based components to improve ease of implementation and enthusiasm of teachers [228]. The Klisch et al's (2013) study, which used interactive game sessions in 11th and 12th graders, found the intervention to be effective in promoting healthier attitudes toward nonmedical use of prescription drugs [224]. Similarly, the Haug study in 2017, which used a mobile phone–based intervention found that it improved study participation, retention, and improved effectiveness with a statistically significant increase in the life skills and self-management skills and reduction in the number at risk for alcohol use [218]. However, a study conducted by Christoff Ade et al in 2015 could not find conclusive evidence of effectiveness of computer-based intervention among college students for reducing substance use [213].

A total of 4 systematic reviews conducted by Champion et al (2013 and 2016), Wood et al (2014), and Hopson et al (2015) on the use of computer- and internet-based programs found them to be potentially efficacious methods of delivering drug prevention programs. The Champion et al's reviews found greatest effects in relation to drug- and alcohol-related knowledge with persisting effectiveness at 6- and 12-month follow-ups [209,214]. The Wood et al's review, on the other hand, emphasized the need for further research to better understand the value of human contact in health interventions and to determine the optimal levels of professional input [207], whereas the Hopson et al's (2015) review identified computer- and internet-based programs as cost-effective options for reaching more individuals, but on the whole reported mixed findings in terms of the effectiveness over traditional methods [208]. In addition, 2 reviews conducted by Jiang et al (2017) and Kazemi et al (2017) reviewed the use of telephone and mobile technology in substance use prevention and found that although it was a promising means to address substance use, the studies included in the reviews for the most part showed either inconclusive or mixed results in terms of the efficiency and efficacy [233,242].

Discussion

Principal Findings

This review included studies spanning across the globe, with the target population for these studies ranging across varying age groups, race or ethnicity, and gender and having differing study designs and sample sizes. The period for the literature search ranged from 1809 to January 2018, but the search only captured studies going as back as the early 1950s. This could be attributed to the fact that the review only included studies that had digitalized records enlisted on the databases searched. However, it can be said with fair amount of certainty that this review manages to capture majority of the trends in the use of CHIT. The boom in the use of media and CHIT platforms was seen to have occurred post the Second World War, thereby reaffirming the validity of the literature search timeline [243].

This review explored multiple CHIT platforms such as television, radio, films, mass media, computer, CD-ROM, internet, social media, and mobile. It was observed that from 1971 to 1989, film, televison, radio, and mass media were the most commonly used modes of intervention [26-30,217,222,225,226,230-232,236-240], whereas post 2009, there was a greater emphasis on the use of computer- and internet-based interventions [206-215,219-221,223,224,227-229,234,241]. Furthermore, recent years show a growing emphasis toward exploring the role of social media– and mobile phone–based interventions to expand the reach of these prevention programs [216,218,233,235,242].

CHIT-based interventions have been shown to overcome challenges imposed by in-person–delivered intervention strategies such as the need for trained personnel to prepare and deliver intervention programs [218]. In addition, studies that examined the impact of electronic health and mobile health (mHealth) interventions found them to be efficacious and cost-effective, with computer-based interventions being more cost-effective than other preventive measures that are labor intensive and costly such as Life Skills Training Program [244]. In our study, the use of media and technology to prevent substance abuse was also found to have several advantages as a prevention strategy. Technology-based interventions can facilitate rapid dissemination of information and improving knowledge about substance use [220,236,238,240]. They can also reduce intervention variability that may occur with a person-based intervention method [226], improving integrity

of intervention measures. In addition, intervention recipients, especially nonabusers who are ideal candidates for primary prevention, are more likely to depend on the media to gain information and knowledge about substance abuse [228].

Studies based on the family interaction theory and aimed at improving relationship quality among mother-daughter duos and studies that were gender-based were found to be effective in reducing the substance use in both the long and short term [220,223,227,234,235]. Multiple studies included in the review used school-based programs for delivery of Web-based interventions; of these, the climate school studies conducted in Australia not only found evidence that internet-based preventive interventions significantly decreased substance use but also demonstrated that they can concurrently reduce associated risk factors in adolescents [209-212,214]. However, the intervention did not show significant impact in the short-term use of substances; the effects of these interventions were only apparent after 12 months, thereby showing a time-delayed effect, attributed to the time required by the students to experience and implement strategies learnt [209,211,212]. Other studies that explored the effectiveness of school-based programs using Web-based or mobile phone–based interventions also found similar results. The studies that focused on the use of social media and mHealth platforms suggested the growth of research and literature in this domain [216,218,233,235,242].

However, in this study, we also found that technology-based interventions are not a panacea in the prevention of substance abuse. Despite the great number of resources poured into development and implementation of media- and technology-based interventions, earlier and recent studies demonstrated moderate effectiveness of these strategies in changing attitude, and ultimately, behavior of recipients of interventions [210,228,235-239]. Although in cases where people may gain greater and more accurate knowledge in substance abuse and negative consequences ensuing from the use, studies failed to show the changes in terms of decreasing or terminating the use or abuse of substance because of these interventions [228,236,237]. For example, studies that explored the role of film, television, radio, and mass media did not find any conclusive evidence to support the stand-alone effectiveness of these platforms. Most of the studies concluded that these platforms should be used in conjunction with other prevention initiatives [26,28,30,222,225,226,237-240].

This review also shows a gradual shift in the types and use of CHIT platforms over the past few decades. It has slowly moved from mass media–based interventions toward Web-based interventions, and following the current trends is heading toward a greater emphasis on telehealth and mHealth-based interventions. We live in an age in which most people frequently use technology and social media and are acutely aware of the current opioid misuse and substance abuse predicament facing by the United States in general and the world at large [22,245]. In this social context, technology could be useful to reach the general population as well as specific at-risk population and potentially be used to develop more tailored and effective prevention. In particular, because adolescents are frequent and avid users of various types of latest technology, computers and smartphones among other technologies could potentially be powerful tools in the primary prevention of substance abuse [218]. Thus, we are quite confident that future research should be focused more on leveraging the use of current CHIT platforms such as mobiles and social media to enhance the outreach of substance abuse prevention programs.

Limitations

This study has multiple strengths and is unique in its approach to map the changing trends in the use of CHIT for substance abuse prevention. It covers a long period and spans across the globe. By design, it did not capture technological interventions for alcohol prevention, as a stand-alone review on the prevention of alcohol abuse would be more appropriate. It is also seen that there was a gap in the literature between 1990 and 2008, no studies during this period were included, and this could be attributed to the stringent eligibility criteria for this review. For study limitations, the review solely focuses on primary prevention and hence, fails to capture the use of CHIT in secondary prevention and its role in treatment of substance use.

The focus of the study on primary prevention of substance abuse necessitated the exclusion of a large number of studies; however, there is an opportunity to follow on with subsequent studies to fill this gap. For example, by design, the study did not capture the effects and correlation between substance use and comorbidities such as HIV and risky health behaviors and the use of CHIT to either treat or prevent either of the repercussions of these correlations. It did not include studies conducted in special populations such as cancer patients and patients with HIV, AIDS, or other STDs. This review also did not include a large number of studies that explored the use of mobile and social media platforms as vehicles of intervention delivery, as opposed to providing prevention programs. This again could be attributed to the stringent eligibility criteria.

Conclusions

This review shows a gradual shift in the types and use of CHIT platforms over the past few decades for substance abuse prevention. It captures the progression from mass media–based interventions toward Web-based intervention and the current trends that head toward a greater emphasis on telehealth- and mHealth-based interventions while emphasizing the need for further development and study of these interventions. It also highlights the gradual shift in consumer and participant behavior, wherein preferences have slowly moved from face-to-face interactions toward more Web- and technology-based platforms, given the anonymity and the vast outreach that these platforms offer. Studies included in this review found these technologies to be effective and cost-effective in real-world settings and contexts. Taking into account the familiarity and ease of use of these CHIT platforms among adults and youth alike, we now have an opportunity to further leverage these platforms for substance use prevention.

Acknowledgments

The views expressed in this report are those of the authors alone. The authors would like to thank the librarian at the University of Maryland, School of Pharmacy, for her assistance with the literature search.

Authors' Contributions

FTS, AMP, and JF conceived the study and contributed to the design of the scoping review. AP and LP contributed to the execution and analysis of the scoping review. FTS and JF contributed significant intellectual content presented in this paper as expert members on the team. All authors were involved in writing the paper. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms used.

[PDF File (Adobe PDF File), 71KB - jmir_v21i1e11297_app1.pdf]

Multimedia Appendix 2

Summary of articles by study design, CHIT used, and outcomes.

[PDF File (Adobe PDF File), 252KB - jmir_v21i1e11297_app2.pdf]

Multimedia Appendix 3

Details on articles included in the reviews cited in the manuscript.

[PDF File (Adobe PDF File), 264KB - jmir_v21i1e11297_app3.pdf]

References

1. Columbia CASA. Center On Addiction. 2012. Addiction medicine: closing the gap between science and practice URL: https://www.centeronaddiction.org/addiction-research/reports/addiction-medicine-closing-gap-between-science-and-practice [WebCite Cache ID 70qGo7G8C]
2. National Institute on Drug Abuse. 2017. Trends and Statistics URL: https://www.drugabuse.gov/related-topics/trends-statistics [accessed 2018-07-11] [WebCite Cache ID 70qH99whJ]
3. American Addiction Centers. 2018. Statistics on Drug Addiction URL: https://americanaddictioncenters.org/rehab-guide/addiction-statistics/ [accessed 2018-07-12] [WebCite Cache ID 70qVoKjh9]
4. 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] [Medline: 19560604]
5. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. 2014. Best Practices for Comprehensive Tobacco Control Programs Atlanta URL: https://www.cdc.gov/tobacco/stateandcommunity/best_practices/pdfs/2014/comprehensive.pdf [WebCite Cache ID 70xUULp4q]
6. National Drug Intelligence Center. United States Department of Jusitce. 2011. The Economic Impact of Illicit Drug Use on American Society URL: http://www.simeoneassociates.com/assets/pdfs/SAI_Assessment1.pdf [accessed 2018-07-17] [WebCite Cache ID 70yybZm6W]
7. Gianluca QG. European Parliament. 2017. Technological innovation strategies in substance use disorders URL: http://www.europarl.europa.eu/RegData/etudes/STUD/2017/581942/EPRS_STU(2017)581942_EN.pdf [WebCite Cache ID 70rJTntzC]
8. Abuse N. National Institute of Drug Abuse. 2017. Overdose Death Rates URL: https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates [accessed 2018-07-12] [WebCite Cache ID 70rZwZgUW]
9. Sloboda Z, Bukoski WJ. The Emerging Science of Drug Abuse Prevention. In: Handbook of drug abuse prevention. Boston: Springer; 2006.
10. National Institute on Alcohol Abuse and Alcoholism. National Institute of Health. 2005. Preventing Alcohol Abuse and Dependence URL: https://pubs.niaaa.nih.gov/publications/social/module3prevention/module3.html [accessed 2018-10-04] [WebCite Cache ID 72vll5lxG]
11. Medina-Mora M. Prevention of substance abuse: a brief overview. World Psychiatry 2005 Feb;4(1):25-30 [FREE Full text] [Medline: 16633497]
12. Substance Abuse and Mental Health Services Administration. 2018. Prevention of Substance Abuse and Mental Illness URL: https://www.samhsa.gov/prevention [accessed 2018-10-05] [WebCite Cache ID 72vmdenrG]

13. Gibbons M, Wilson R, Samal L, Lehmann C, Dickersin K, Lehmann H, et al. Consumer health informatics: results of a systematic evidence review and evidence based recommendations. Transl Behav Med 2011 Mar;1(1):72-82 [FREE Full text] [doi: 10.1007/s13142-011-0016-4] [Medline: 24073033]
14. Or CK, Karsh B, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. J Am Med Inform Assoc 2011 Feb;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
15. Gustafson DH, Hawkins RP, Boberg EW, McTavish F, Owens B, Wise M, et al. CHESS: 10 years of research and development in consumer health informatics for broad populations, including the underserved. Int J Med Inform 2002 Nov 12;65(3):169-177. [Medline: 12414016]
16. Lewis D, Eysenbach G, Kukafka R. Consumer Health Informatics: Informing Consumers and Improving Health Care. New York: Springer-Verlag; 2005.
17. Tao D, Wang T, Wang T, Liu S, Qu X. Effects of consumer-oriented health information technologies in diabetes management over time: a systematic review and meta-analysis of randomized controlled trials. J Am Med Inform Assoc 2017 Sep 01;24(5):1014-1023. [doi: 10.1093/jamia/ocx014] [Medline: 28340030]
18. Or C, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
19. Finkelstein J, Knight A, Marinopoulos S, Gibbons MC, Berger Z, Aboumatar H, et al. Enabling patient-centered care through health information technology. Evid Rep Technol Assess (Full Rep) 2012 Jun(206):1-1531. [Medline: 24422882]
20. PEW Research Center. 2018. Demographics of Internet and Home Broadband Usage in the United States URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [WebCite Cache ID 70raT3uGu]
21. PEW Research Center. Three Technology Revolutions URL: http://www.pewinternet.org/three-technology-revolutions/ [accessed 2018-07-12] [WebCite Cache ID 70rawUWun]
22. World Health Organisation. E-health technologies and substance abuse URL: http://www.who.int/substance_abuse/activities/ehealth/en/ [WebCite Cache ID 70rbPY4gk]
23. Marsch LA, Borodovsky JT. Technology-based interventions for preventing and treating substance use among youth. Child Adolesc Psychiatr Clin N Am 2016 Dec;25(4):755-768. [doi: 10.1016/j.chc.2016.06.005] [Medline: 27613350]
24. 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]
25. Young SD, Swendeman D, Holloway IW, Reback CJ, Kao U. Use of technology to address substance use in the context of HIV: a systematic review. Curr HIV/AIDS Rep 2015 Dec;12(4):462-471 [FREE Full text] [doi: 10.1007/s11904-015-0295-3] [Medline: 26475670]
26. Feingold PC, Knapp ML. Anti-drug abuse commercials. J Commun 1977;27(1):20-28. [Medline: 833346]
27. Miller MC, Cantor AB, Larisey L, Murphy E. Comparison of media for substance abuse education in rural communities. Int J Addict 1981 May;16(4):585-592. [Medline: 7287242]
28. Milne HB, Butt TW. The critical assessment of the T.V. film “The Drug Takers”. Br J Addict Alcohol Other Drugs 1975 Jun;70(2):216-220. [Medline: 1058026]
29. Pickens KA. The use of films in drug education--a review. Health Educ J 1984;43(4):19-20. [Medline: 10317570]
30. Sussman S, Flay B, Sobel J, Rauch J, Hansen W, Johnson C. Viewing and evaluation of a televised drug education program by students previously or concurrently exposed to school-based substance abuse prevention programming. Health Educ Res 1987;2(4):373-383. [doi: 10.1093/her/2.4.373]
31. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
32. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
33. Faulkner GE, Grootendorst P, Nguyen VH, Andreyeva T, Arbour-Nicitopoulos K, Auld MC, et al. Economic instruments for obesity prevention: results of a scoping review and modified Delphi survey. Int J Behav Nutr Phys Act 2011 Oct 06;8:109 [FREE Full text] [doi: 10.1186/1479-5868-8-109] [Medline: 21978599]
34. Bottorff JL, Haines-Saah R, Kelly MT, Oliffe JL, Torchalla I, Poole N, et al. Gender, smoking and tobacco reduction and cessation: a scoping review. Int J Equity Health 2014 Dec 12;13:114 [FREE Full text] [doi: 10.1186/s12939-014-0114-2] [Medline: 25495141]
35. Kanter D. The drug scene: current research. Public Opin Q 1971 Sep 01;35(3):459-463. [doi: 10.1093/poq/35.3.459]
36. O'Keefe M. The Anti-Smoking Commercials: A Study of Television's Impact on Behavio. Public Opin Q 1971;35(2):242-248 [FREE Full text]
37. Fejer D, Smart R, Whitehead P, Laforest L. Sources of information about drugs among high school students. Public Opin Q 1971;35(2):235-241. [doi: 10.1086/267895]
38. Hanneman G. Sources of Drug Abuse Information on the College Campus. Connecticut: University of Connecticut; 1972.
39. Linsky AS. The changing public views of alcoholism. Q J Stud Alcohol 1970 Sep;31(3):692-704. [Medline: 5466398]
40. Lipp M, Benson S, Taintor Z. Marijuana use by medical students. Am J Psychiatry 1971 Aug;128(2):207-212. [doi: 10.1176/ajp.128.2.207] [Medline: 5112947]

41. Pollock SH. Attitudes of medical students toward marijuana. J Psychoactive Drugs 2012 Jan 17;5(1):56-61. [doi: 10.1080/02791072.1972.10471470]
42. Grant J. Drug education based on a knowledge, attitude, and experience study. J Sch Health 1971 Sep;41(7):383-386. [Medline: 5208626]
43. Zajonc RB. Attitudinal effects of mere exposure. J Pers Soc Psychol 1968;9(2, Pt.2):1-27. [doi: 10.1037/h0025848]
44. Amendolara F. Modifying attitudes towards drugs in seventh grade students. J Drug Educ 1995 Jan;3(1):71-78. [doi: 10.2190/WL56-DEV7-441M-LDDW]
45. Braxton ER, Yonker RJ. Does being urban, poor, black, or female affect youth's knowledge and-or attitudes relating to drugs? J Sch Health 1973 Mar;43(3):185-188. [Medline: 4486608]
46. Einstein S, Lavehar M, Garitano WW. Drug abuse education and the multiplier effect: an experience in training 109 teachers. J Sch Health 1972 Dec;42(10):609-613. [Medline: 4485223]
47. Einstein S, Lavenhar MA, Wolfson EA, Louria DB, Quinones MA, McAteer G. The training of teachers for drug abuse education programs: preliminary considerations. J Drug Educ 1995 Jan;1(4):323-345. [doi: 10.2190/12A7-WQR7-1B27-J2DF]
48. Richardson DW, Nader PR, Roghmann KJ, Friedman S. Attitudes of fifth grade students to illicit psychoactive drugs. J Sch Health 1972 Sep;42(7):389-391. [Medline: 4483996]
49. Irwin R, Creswell W, Stauffer D. The effect of the teacher and three different Classroom approaches on seventh grade students? Knowledge, attitudes and beliefs about smoking*. J Sch Health 1970;40(7):355-359. [doi: 10.1111/j.1746-1561.1970.tb05622.x]
50. Levitt L, Baganz P, Blachly P. A study of employee attitudes toward patients in a hospital for the treatment of drug addiction. Psychiatr Q 1963 Apr;37:210-219. [Medline: 13930213]
51. Shaw CT. Knowledge and attitude responses of college students toward controversial social health issues. J Sch Health 1972 Jan;42(1):53-56. [Medline: 4480910]
52. Barcus F. Drug advertising on television. In: Drug Use in America: Problem in Perspective. Washington, DC: Irvington Pub; 1973.
53. Hanneman G, Eisenstock B, Weinbeck W. The medicine man message: an evaluation of a California state office of narcotics and drug abuse prevention campaign to inform the public of the dangers of prescription and over-the -counter drug misuse. California: University of Southern California, Pacific Institute for Research and Evaluation; 1977.
54. Warner KE. The effects of the anti-smoking campaign on cigarette consumption. Am J Public Health 1977 Jul;67(7):645-650. [Medline: 879393]
55. Wotring C, Heald G, Carpenter C. Evaluation of the Florida drug abuse campaign (1976-1977). Florida state university: College of communication; 1977.
56. Wallack L. An assessment of drinking patterns, problems., knowledge and attidtudes in three Northern California communities. Social research group 1978;Contract No:C21.
57. Brecher EM. Licit and Illicit Drugs: The Consumers Union Report on Narcotics, Stimulants, Depressants, Inhalants, Hallucinogens, and Marijuana - Including Caffeine, Nicotine, and Alcohol. United States: Little, Brown and Company; 1973.
58. Lazarsfeld P, Merton RK. Mass communication, popular taste,organized social aciton. In: Schramm W, Roberts D, editors. The process and effects of mass communication. Revised edition. United States: University of Illinois Press; 1971.
59. Fishbein F. Consumer beliefs and behavior with respect to cigarette smoking : a critical analysis of the public literature. A report prepared for the staff of the Federal Trade Commission. Champaign, Ill: University of Illinois at Champaign-Urbana; 1977.
60. Department of national health and welfare. Long range health planning b, non-medical use of d. Smoking and health in Canada 1977.
61. Hanneman G, McEwen W. Televised drug abuse appeals: a content analysis. Journal Mass Commun Q 2016 Aug 29;50(2):329-333. [doi: 10.1177/107769907305000216]
62. Schlegel R. The role of persuasive communications in drug dissuasion. J Drug Educ 1995 Jan;7(3):279-290. [doi: 10.2190/XWTJ-X7CT-0PGD-Y70E]
63. Dembo R, Schmeidler J, Lipton DS, Babst DV, Diamond SC, Spielman CR, et al. A survey of students' awareness of and attitudes toward drug abuse prevention programs in New York State, winter 1974/75. Int J Addict 1979 Apr;14(3):311-328. [Medline: 468427]
64. Kline JA. Evaluation of a Multimedia Drug Education Program. J Drug Educ 1995 Jan;2(3):229-239. [doi: 10.2190/NW0E-8FQU-N6RN-68QP]
65. Dembo R, Miran M, Babst DV, Schmeidler J. The believability of the media as sources of information on drugs. Int J Addict 1977 Oct;12(7):959-969. [Medline: 591149]
66. Smart R, Fejer D. Credibility of sources of drug information for high school students. J Drug Issues 1972 Apr;2(2):8-18. [doi: 10.1177/002204267200200204]
67. Milavsky J, Pekowsky B, Stipp H. TV drug advertising and proprietary and illicit drug use among teenage boys*. Public Opin Q 1975;39(4):457-481. [doi: 10.1086/268244]

68. Payne D. The Relationship between Television Advertising and Drug Abuse among Youth: Fancy and Fact. New York: Baywood Publishing Co; 1976:215-220.
69. Kohn P, Snook S. Balanced vs. one-sided communications about drugs. J Drug Educ 1995 Jan;6(4):273-281. [doi: 10.2190/M3LL-K013-VRUY-WLGP]
70. Wotring C, Heald G, Carpenter C, Schmeling D. Attacking the drug norm: effects of the 1976–77 Florida Drug Abuse TV Campaign. J Drug Educ 1995 Jan 25;9(3):255-261. [doi: 10.2190/1AFQ-GHM7-JL7U-U7VA]
71. Schmeling D, Wotring C. Agenda-setting effects of drug abuse public service ads. J Q 2016 Aug 29;53(4):743-746. [doi: 10.1177/107769907605300424]
72. Hanneman G. The medicine man message: a delayed effects evaluation of a California state office of narcotics and drug abuse prevention campaign to inform the public of the dangers of prescription and over-the -counter drug misuse. In: University of Southern California, Center for communications policy research. United States: University of Southern California; 1978:a.
73. Sandmaier M. Sandmaier M. Myths and messages: using the media as a prevention tool, in the women next door: summary proceedings of a symposium on the subject of drugs and the modern women. US journal of drug and alcohol dependence 1980:a.
74. Capalaces R, Starr J. The negative message of anti-drug spots: does it get across? Public Telecommunications Review 1973;1:64-66.
75. Delaney R. Florida study looks at effects of media messages. National Institute of Alcohol Abuse and Alcoholism information and feature service 1981:5.
76. Dickman FB, Keil TJ. Public television and public health. The case of alcoholism. J Stud Alcohol 1977 Mar;38(3):584-592. [Medline: 859338]
77. Field T, Deitrick S, Hersey J, Probst J, Theologus G. Implementing public education campaigns:lessons from alcohol abuse prevention. Summary report to NIAAA. Washington, DC: Kappa systems; 1983.
78. Hanneman GJ, McEwen WJ, Coyne SA. Public service advertising on television. Journal of Broadcasting 1973 Sep;17(4):387-404. [doi: 10.1080/08838157309363703]
79. Morrison A, Kline F, Miller P. Aspects of adolescent information acquisition about drugs alcohol topics. In: Ostman R, editor. Communication Research and Drug Education. London: Sage publication; 1976.
80. Plant MA, Pirie F, Kreitman N. Evaluation of the Scottish Health Education Unit's 1976 campaign on alcoholism. Soc Psychiatry 1979;14(1):11-24. [doi: 10.1007/BF00583569]
81. Rappeport M, Labow P. The public evaluates the NIAAA public education campaign: a sudy for the U.S. department of health, education, welfare, public health service, alcohol, drug abuse, and mental health administration. Princeton, New Jersey: Opinion research corporation; 1975.
82. Trager R. Adolescent reactions to educational media messages regarding drug education. In: Ostman R, editor. Communication research and drug education. London: Sage Publications; 1976.
83. Wong M, Barbatsis G. Attitude and information change effected by drug education via broadcast television and group viewing. J Drug Education 1976;8(2):161.
84. Weimer J. The effects of film treatments on attitudes that correlated with drug-behavior. South Dakota: University of South Dakota; 1976.
85. English GE. The effectiveness of emotional appeal versus fact-giving drug educational films. J Sch Health 1972;42(9):540-541. [doi: 10.1111/j.1746- 1561.1972.tb00975.x]
86. Swift B, Dorn N, Thompson A. Evaluation of drug education : findings of a national research study of effects on secondary school students of five types of lessons given by teachers. London: Institute for the study of drug dependence; 1974.
87. Thornell J. The construction and evalution of drug education of a drug education programme for third grade students. Boston, MA: Boston College; 1976.
88. Sohn M. Change in factual knowledge and reported use of illicit drugs resulting from the viewing of a motion picture. United States: University of Maryland; 1976.
89. Taussig W. The effects of a family life program and a drug education program on the self esteem of fifth grade children. New York: New York University; 1978.
90. Rosengren K, Wenner L, Palmgreen P. Media Gratifications Research: Current Perspectives. London: SAGE Publications; 1985.
91. Flay B, Pentz M, Johnson C, Sussman S. Reaching children with mass media health promotion programs: the relative effectivenss of an advertising campaign, a community-based programa school-based program. In: Leather G, editor. Health evaluation and the media. Pergamon: Journal of School Health; 1986.
92. Flay B. Psychosocial approaches to smoking prevention: a review of findings. Health Psychol 1985;4(5):449-488. [Medline: 3908090]
93. Flay BR. Mass media and smoking cessation: a critical review. Am J Public Health 1987 Feb;77(2):153-160. [Medline: 3541650]
94. National Coalition of Hispanic Health and Human Services Organization. Early intervention with Hispanic youth. Washington, DC: Office for Substance Abuse Prevention, Alcohol, Drug Abuse and Mental Health Administration. (COSSMHO); 1988.

95. Amaro H, Campa R, Coffman G, Heeren T. Initiation of substance abuse among Mexican American, Cuban American, and Puerto Rican adolescents and young adults:findings from the Hispanic HANES. Am J Public Health 1989.
96. Booth M, Castro F, Anglin M. What do we know about Hispanic substance abuse? a review of the literature. In: Glick R, Moore J, editors. Drugs in Hispanic communities. New Brunswick, NJ: Rutgers University Press; 1989.
97. Caetano R, Martinez R. Alcohol use in Madrid and among U.S. Hispanics. Berkeley, CA: The National Institute on Alcohol Abuse and Alcoholism; 1987.
98. Buller DB, Borland R, Woodall WG, Hall JR, Hines JM, Burris-Woodall P, et al. Randomized trials on consider this, a tailored, internet-delivered smoking prevention program for adolescents. Health Educ Behav 2008 Apr;35(2):260-281 [FREE Full text] [doi: 10.1177/1090198106288982] [Medline: 17114331]
99. Norman CD, Maley O, Li X, Skinner HA. Using the internet to assist smoking prevention and cessation in schools: a randomized, controlled trial. Health Psychol 2008 Nov;27(6):799-810. [doi: 10.1037/a0013105] [Medline: 19025276]
100. Prokhorov AV, Kelder SH, Shegog R, Murray N, Peters Jr R, Agurcia-Parker C, et al. Impact of A Smoking Prevention Interactive Experience (ASPIRE), an interactive, multimedia smoking prevention and cessation curriculum for culturally diverse high-school students. Nicotine Tob Res 2008 Sep;10(9):1477-1485. [doi: 10.1080/14622200802323183] [Medline: 19023839]
101. Vogl L, Teesson M, Andrews G, Bird K, Steadman B, Dillon P. A computerized harm minimization prevention program for alcohol misuse and related harms: randomized controlled trial. Addiction 2009 Apr;104(4):564-575. [doi: 10.1111/j.1360-0443.2009.02510.x] [Medline: 19335655]
102. Newton NC, Vogl LE, Teesson M, Andrews G. CLIMATE Schools: alcohol module: cross-validation of a school-based prevention programme for alcohol misuse. Aust N Z J Psychiatry 2009 Mar;43(3):201-207. [doi: 10.1080/00048670802653364] [Medline: 19221908]
103. Newton NC, Andrews G, Teesson M, Vogl LE. Delivering prevention for alcohol and cannabis using the Internet: a cluster randomised controlled trial. Prev Med 2009 Jun;48(6):579-584. [doi: 10.1016/j.ypmed.2009.04.009] [Medline: 19389420]
104. Newton NC, Teesson M, Vogl LE, Andrews G. Internet-based prevention for alcohol and cannabis use: final results of the Climate Schools course. Addiction 2010 Apr;105(4):749-759. [doi: 10.1111/j.1360-0443.2009.02853.x] [Medline: 20148791]
105. Koning IM, Vollebergh WA, Smit F, Verdurmen JEE, Van Den Eijnden RJ, Ter Bogt TF, et al. Preventing heavy alcohol use in adolescents (PAS): cluster randomized trial of a parent and student intervention offered separately and simultaneously. Addiction 2009 Oct;104(10):1669-1678. [doi: 10.1111/j.1360-0443.2009.02677.x] [Medline: 21265908]
106. Aveyard P, Sherratt E, Almond J, Lawrence T, Lancashire R, Griffin C, et al. The change-in-stage and updated smoking status results from a cluster-randomized trial of smoking prevention and cessation using the transtheoretical model among British adolescents. Prev Med 2001 Oct;33(4):313-324. [doi: 10.1006/pmed.2001.0889] [Medline: 11570836]
107. Marsch LA, Bickel WK, Grabinski MJ. Application of interactive, computer technology to adolescent substance abuse prevention and treatment. Adolesc Med State Art Rev 2007 Aug;18(2):342-56, xii. [Medline: 18605650]
108. Duncan TE, Duncan SC, Beauchamp N, Wells J, Ary DV. Development and evaluation of an interactive CD-ROM refusal skills program to prevent youth substance use: “refuse to use”. J Behav Med 2000 Feb;23(1):59-72. [Medline: 10749011]
109. Lord S, D’Amante D. Efficacy of online alcohol and other drug prevention for early adolescents. J Adolesc Health 2007 Feb;40(2):S4. [doi: 10.1016/j.jadohealth.2006.11.016]
110. Budney AJ, Fearer S, Walker DD, Stanger C, Thostenson J, Grabinski M, et al. An initial trial of a computerized behavioral intervention for cannabis use disorder. Drug Alcohol Depend 2011 May 1;115(1-2):74-79 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.10.014] [Medline: 21131143]
111. Gilbert P, Ciccarone D, Gansky SA, Bangsberg DR, Clanon K, McPhee SJ, et al. Interactive “Video Doctor” counseling reduces drug and sexual risk behaviors among HIV-positive patients in diverse outpatient settings. PLoS One 2008 Apr 23;3(4):e1988 [FREE Full text] [doi: 10.1371/journal.pone.0001988] [Medline: 18431475]
112. Kay-Lambkin FJ, Baker AL, Lewin TJ, Carr VJ. Computer-based psychological treatment for comorbid depression and problematic alcohol and/or cannabis use: a randomized controlled trial of clinical efficacy. Addiction 2009 Mar;104(3):378-388. [doi: 10.1111/j.1360-0443.2008.02444.x] [Medline: 19207345]
113. Lee CM, Neighbors C, Kilmer JR, Larimer ME. A brief, web-based personalized feedback selective intervention for college student marijuana use: a randomized clinical trial. Psychol Addict Behav 2010 Jun;24(2):265-273 [FREE Full text] [doi: 10.1037/a0018859] [Medline: 20565152]
114. Tossmann H, Jonas B, Tensil M, Lang P, Strüber E. A controlled trial of an internet-based intervention program for cannabis users. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):673-679. [doi: 10.1089/cyber.2010.0506] [Medline: 21651419]
115. Williams C, Griffin KW, Macaulay AP, West TL, Gronewold E. Efficacy of a drug prevention CD-ROM intervention for adolescents. Subst Use Misuse 2005;40(6):869-878. [Medline: 15974146]
116. Bersamin M, Paschall MJ, Fearnow-Kenney M, Wyrick D. Effectiveness of a web-based alcohol-misuse and harm-prevention course among high- and low-risk students. J Am Coll Health 2007;55(4):247-254. [doi: 10.3200/JACH.55.4.247-254] [Medline: 17319331]
117. Bingham CR, Barretto AI, Walton MA, Bryant CM, Shope JT, Raghunathan TE. Efficacy of a web-based, tailored, alcohol prevention/intervention program for college students: 3-month follow-up. J Drug Educ 2011;41(4):405-430 [FREE Full text] [Medline: 22455103]

118. Bishop D, Bryant KS, Giles SM, Hansen WB, Dusenbury L. Simplifying the delivery of a prevention program with web-based enhancements. J Prim Prev 2006 Jul;27(4):433-444. [doi: 10.1007/s10935-006-0042-z] [Medline: 16763766]
119. Croom K, Lewis D, Marchell T, Lesser ML, Reyna VF, Kubicki-Bedford L, et al. Impact of an online alcohol education course on behavior and harm for incoming first-year college students: short-term evaluation of a randomized trial. J Am Coll Health 2009;57(4):445-454. [doi: 10.3200/JACH.57.4.445-454] [Medline: 19114384]
120. Di Noia J, Schwinn TM, Dastur ZA, Schinke SP. The relative efficacy of pamphlets, CD-ROM, and the Internet for disseminating adolescent drug abuse prevention programs: an exploratory study. Prev Med 2003 Dec;37(6 Pt 1):646-653 [FREE Full text] [Medline: 14636798]
121. Epstein J, Collins K, Thomson N, Pancella T, Pauley D. The doubles: evaluation of a substance abuse education curriculum for elementary school students. J Child Adolesc Subst Abuse 2007;16(4):1-22. [doi: 10.1300/J029v16n04_01]
122. Hecht ML, Marsiglia FF, Elek E, Wagstaff DA, Kulis S, Dustman P, et al. Culturally grounded substance use prevention: an evaluation of the keepin' it R.E.A.L. curriculum. Prev Sci 2003 Dec;4(4):233-248. [Medline: 14598996]
123. Hustad JT, Barnett NP, Borsari B, Jackson KM. Web-based alcohol prevention for incoming college students: a randomized controlled trial. Addict Behav 2010 Mar;35(3):183-189 [FREE Full text] [doi: 10.1016/j.addbeh.2009.10.012] [Medline: 19900763]
124. Koning IM, van den Eijnden RJ, Verdurmen JE, Engels RC, Vollebergh WA. Long-term effects of a parent and student intervention on alcohol use in adolescents: a cluster randomized controlled trial. Am J Prev Med 2011 May;40(5):541-547. [doi: 10.1016/j.amepre.2010.12.030] [Medline: 21496753]
125. Kypri K, Saunders JB, Williams SM, McGee RO, Langley JD, Cashell-Smith ML, et al. Web-based screening and brief intervention for hazardous drinking: a double-blind randomized controlled trial. Addiction 2004 Nov;99(11):1410-1417. [doi: 10.1111/j.1360-0443.2004.00847.x] [Medline: 15500594]
126. Moore MJ, Soderquist J, Werch C. Feasibility and efficacy of a binge drinking prevention intervention for college students delivered via the Internet versus postal mail. J Am Coll Health 2005;54(1):38-44. [doi: 10.3200/JACH.54.1.38-44] [Medline: 16050327]
127. Neighbors C, Lewis MA, Atkins DC, Jensen MM, Walter T, Fossos N, et al. Efficacy of web-based personalized normative feedback: a two-year randomized controlled trial. J Consult Clin Psychol 2010 Dec;78(6):898-911 [FREE Full text] [doi: 10.1037/a0020766] [Medline: 20873892]
128. 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] [Medline: 19757052]
129. Schinke SP, Cole KC, Fang L. Gender-specific intervention to reduce underage drinking among early adolescent girls: a test of a computer-mediated, mother-daughter program. J Stud Alcohol Drugs 2009 Jan;70(1):70-77 [FREE Full text] [Medline: 19118394]
130. Schinke SP, Schwinn TM, Di Noia J, Cole KC. Reducing the risks of alcohol use among urban youth: three-year effects of a computer-based intervention with and without parent involvement. J Stud Alcohol 2004 Jul;65(4):443-449 [FREE Full text] [Medline: 15376818]
131. Walters ST, Vader AM, Harris TR. A controlled trial of web-based feedback for heavy drinking college students. Prev Sci 2007 Mar;8(1):83-88. [doi: 10.1007/s11121-006-0059-9] [Medline: 17136461]
132. Warren J, Hecht M, Wagstaff D, Elek E, Ndiaye K, Dustman P, et al. Communicating prevention: the effects of the keepin' it REAL classroom videotapes and televised PSAs on middle-school students' substance use. J Appl Commun Res 2006 May;34(2):209-227. [doi: 10.1080/00909880600574153]
133. de Josselin de Jong JS, Candel M, Segaar D, Cremers H, de Vries H. Efficacy of a web-based computer-tailored smoking prevention intervention for Dutch adolescents: randomized controlled trial. J Med Internet Res 2014;16(3):e82 [FREE Full text] [doi: 10.2196/jmir.2469] [Medline: 24657434]
134. Doumas DM, Esp S, Turrisi R, Hausheer R, Cuffee C. A test of the efficacy of a brief, web-based personalized feedback intervention to reduce drinking among 9th grade students. Addict Behav 2014 Jan;39(1):231-238. [doi: 10.1016/j.addbeh.2013.10.011] [Medline: 24148137]
135. Doumas DM, Hausheer R, Esp S, Cuffee C. Reducing alcohol use among 9th grade students: 6 month outcomes of a brief, Web-based intervention. J Subst Abuse Treat 2014 Jul;47(1):102-105. [doi: 10.1016/j.jsat.2014.02.006] [Medline: 24666810]
136. Vogl LE, Newton NC, Champion KE, Teesson M. A universal harm-minimisation approach to preventing psychostimulant and cannabis use in adolescents: a cluster randomised controlled trial. Subst Abuse Treat Prev Policy 2014 Jun 18;9:24 [FREE Full text] [doi: 10.1186/1747-597X-9-24] [Medline: 24943829]
137. Malmberg M, Kleinjan M, Overbeek G, Vermulst A, Lammers J, Monshouwer K, et al. Substance use outcomes in the Healthy School and Drugs program: results from a latent growth curve approach. Addict Behav 2015 Mar;42:194-202. [doi: 10.1016/j.addbeh.2014.11.021] [Medline: 25481454]
138. Malmberg M, Kleinjan M, Overbeek G, Vermulst A, Monshouwer K, Lammers J, et al. Effectiveness of the 'Healthy School and Drugs' prevention programme on adolescents' substance use: a randomized clustered trial. Addiction 2014 Jun;109(6):1031-1040. [doi: 10.1111/add.12526] [Medline: 24612164]

139. Velicer WF, Redding CA, Paiva AL, Mauriello LM, Blissmer B, Oatley K, et al. Multiple behavior interventions to prevent substance abuse and increase energy balance behaviors in middle school students. Transl Behav Med 2013 Mar;3(1):82-93 [FREE Full text] [doi: 10.1007/s13142-013-0197-0] [Medline: 23585821]
140. Bannink R, Broeren S, Joosten-van Zwanenburg E, van As E, van de Looij-Jansen P, Raat H. Effectiveness of a Web-based tailored intervention (E-health4Uth) and consultation to promote adolescents' health: randomized controlled trial. J Med Internet Res 2014 May 30;16(5):e143 [FREE Full text] [doi: 10.2196/jmir.3163] [Medline: 24878521]
141. Rundle-Thiele S, Schuster L, Dietrich T, Russell-Bennett R, Drennan J, Leo C, et al. Maintaining or changing a drinking behavior? GOKA's short-term outcomes. J Bus Res 2015 Oct;68(10):2155-2163. [doi: 10.1016/j.jbusres.2015.03.015]
142. Walton MA, Resko S, Barry KL, Chermack ST, Zucker RA, Zimmerman MA, et al. A randomized controlled trial testing the efficacy of a brief cannabis universal prevention program among adolescents in primary care. Addiction 2014 May;109(5):786-797 [FREE Full text] [doi: 10.1111/add.12469] [Medline: 24372937]
143. Bauman KE, LaPrelle J, Brown JD, Koch GG, Padgett CA. The influence of three mass media campaigns on variables related to adolescent cigarette smoking: results of a field experiment. Am J Public Health 1991 May;81(5):597-604. [Medline: 2014859]
144. Bauman KE, Brown JD, Bryan ES, Fisher LA, Padgett CA, Sweeney JM. Three mass media campaigns to prevent adolescent cigarette smoking. Prev Med 1988 Sep;17(5):510-530. [Medline: 3237655]
145. Bauman K, Padgett C, Koch G. A media-based campaign to encourage personal communication among adolescents about not smoking cigarettes: participation, selection and consequences. Health Educ Res 1989;4(1):35-44. [doi: 10.1093/her/4.1.35]
146. Brown JD, Bauman KE, Padgett CA. A validity problem in measuring exposure to mass media campaigns. Health Educ Q 1990;17(3):299-306. [Medline: 2228632]
147. L J, Bauman KE, Koch GG. High intercommunity variation in adolescent cigarette smoking in a 10-community field experiment. Eval Rev 2016 Jul 26;16(2):115-130. [doi: 10.1177/0193841x9201600201]
148. Fallin A, Neilands TB, Jordan JW, Hong JS, Ling PM. Wreaking “havoc” on smoking: social branding to reach young adult “partiers” in Oklahoma. Am J Prev Med 2015 Jan;48(1 Suppl 1):S78-S85 [FREE Full text] [doi: 10.1016/j.amepre.2014.09.008] [Medline: 25528713]
149. Flay BR, Miller TQ, Hedeker D, Siddiqui O, Britton CF, Brannon BR, et al. The television, school, and family smoking prevention and cessation project. VIII. Student outcomes and mediating variables. Prev Med 1995 Jan;24(1):29-40. [doi: 10.1006/pmed.1995.1005] [Medline: 7740012]
150. Flay BR, Brannon BR, Johnson CA, Hansen WB, Ulene AL, Whitney-Saltiel DA, et al. The television school and family smoking prevention and cessation project. 1. Theoretical basis and program development. Prev Med 1988 Sep;17(5):585-607. [Medline: 3237658]
151. Sussman S, Brannon BR, Flay BR, Gleason L, Senor S, Sobol DF, et al. The television, school and family smoking prevention/cessation project. II. Formative evaluation of television segments by teenagers and parents – implications for parental involvement in drug education. Health Educ Res 1986;1(3):185-194. [doi: 10.1093/her/1.3.185]
152. Sussman S, Dent CW, Brannon BR, Glowacz K, Gleason LR, Ullery S, et al. The television, school and family smoking prevention/cessation project. IV. Controlling for program success expectancies across experimental and control conditions. Addict Behav 1989;14(6):601-610. [Medline: 2618843]
153. Flynn BS, Worden JK, Secker-Walker RH, Badger GJ, Geller BM. Cigarette smoking prevention effects of mass media and school interventions targeted to gender and age groups. Journal of Health Education 2013 Mar 08;26(sup2):S45-S51. [doi: 10.1080/10556699.1995.10603147]
154. Flynn BS, Worden JK, Secker-Walker RH, Badger GJ, Geller BM, Costanza MC. Prevention of cigarette smoking through mass media intervention and school programs. Am J Public Health 1992 Jun;82(6):827-834. [Medline: 1585963]
155. Flynn BS, Worden JK, Secker-Walker RH, Pirie PL, Badger GJ, Carpenter JH. Long-term responses of higher and lower risk youths to smoking prevention interventions. Prev Med 1997;26(3):389-394. [doi: 10.1006/pmed.1997.0159] [Medline: 9144764]
156. Flynn BS, Worden JK, Secker-Walker RH, Pirie PL, Badger GJ, Carpenter JH, et al. Mass media and school interventions for cigarette smoking prevention: effects 2 years after completion. Am J Public Health 1994 Jul;84(7):1148-1150. [Medline: 8017542]
157. Secker-Walker RH, Worden JK, Holland RR, Flynn BS, Detsky AS. A mass media programme to prevent smoking among adolescents: costs and cost effectiveness. Tob Control 1997;6(3):207-212 [FREE Full text] [Medline: 9396105]
158. Worden JK, Flynn BS. Effective use of mass media to prevent cigarette smoking. J Public Health Manag Pract 2000 May;6(3):vii-viii. [Medline: 10848475]
159. Worden JK, Flynn BS, Geller BM, Chen M, Shelton LG, Secker-Walker RH, et al. Development of a smoking prevention mass media program using diagnostic and formative research. Prev Med 1988 Sep;17(5):531-558. [Medline: 3237656]
160. Worden JK, Flynn BS, Solomon LJ, Secker-Walker RH, Badger GJ, Carpenter JH. Using mass media to prevent cigarette smoking among adolescent girls. Health Educ Q 1996 Nov;23(4):453-468. [Medline: 8910024]
161. Flynn BS, Worden JK, Bunn JY, Solomon LJ, Ashikaga T, Connolly SW, et al. Mass media interventions to reduce youth smoking prevalence. Am J Prev Med 2010 Jul;39(1):53-62 [FREE Full text] [doi: 10.1016/j.amepre.2010.03.008] [Medline: 20537841]

162. Solomon LJ, Bunn JY, Flynn BS, Pirie PL, Worden JK, Ashikaga T. Mass media for smoking cessation in adolescents. Health Educ Behav 2009 Aug;36(4):642-659. [doi: 10.1177/1090198106298421] [Medline: 17602098]
163. Hafstad A, Aaro L. Activating interpersonal influence through provocative appeals: evaluation of a mass media-based antismoking campaign targeting adolescents. Health Commun 1997 Jul;9(3):253-272. [doi: 10.1207/s15327027hc0903_4]
164. Hafstad A. Provocative anti-smoking appeals in mass-media campaigns: an intervention study on adolescent smoking. Norway: Norwegian Women's Public Health Organisation; 1997.
165. Hafstad A, Aarø LE, Engeland A, Andersen A, Langmark F, Stray-Pedersen B. Provocative appeals in anti-smoking mass media campaigns targeting adolescents--the accumulated effect of multiple exposures. Health Educ Res 1997 Jun;12(2):227-236. [Medline: 10168574]
166. Hafstad A. Use of provocative emotional appeals in a mass media campaign designed to prevent smoking among adolescents. Eur J Public Health 1997 Jun 01;7(2):122-127. [doi: 10.1093/eurpub/7.2.122]
167. Longshore D, Ghosh-Dastidar B, Ellickson PL. National Youth Anti-Drug Media Campaign and school-based drug prevention: evidence for a synergistic effect in ALERT Plus. Addict Behav 2006 Mar;31(3):496-508. [doi: 10.1016/j.addbeh.2005.05.032] [Medline: 15979245]
168. Ellickson PL, McCaffrey DF, Ghosh-Dastidar B, Longshore DL. New inroads in preventing adolescent drug use: results from a large-scale trial of project ALERT in middle schools. Am J Public Health 2003 Nov;93(11):1830-1836. [Medline: 14600049]
169. Ghosh-Dastidar B, Longshore DL, Ellickson PL, McCaffrey DF. Modifying pro-drug risk factors in adolescents: results from project ALERT. Health Educ Behav 2004 Jun;31(3):318-334. [doi: 10.1177/1090198104263333] [Medline: 15155043]
170. Longshore D, Ellickson PL, McCaffrey DF, St Clair P. School-based drug prevention among at-risk adolescents: effects of ALERT plus. Health Educ Behav 2007 Aug;34(4):651-668. [doi: 10.1177/1090198106294895] [Medline: 17567823]
171. Worden J, Flynn B. Using television messages to prevent smoking among adolescents. University of Vermont: Office of Health Promotion Research and Biometry Facility, College of Medicine, Department of Mathematics (Statistics Program) College of Engineering and Mathematics; 1983 Presented at: American Public Health Association Annual Meeting; 1983; Dallas.
172. Stotts AL, Diclemente CC, Dolan-Mullen P. One-to-one: a motivational intervention for resistant pregnant smokers. Addict Behav 2002;27(2):275-292. [Medline: 11817768]
173. Rigotti NA, Park ER, Regan S, Chang Y, Perry K, Loudin B, et al. Efficacy of telephone counseling for pregnant smokers: a randomized controlled trial. Obstet Gynecol 2006 Jul;108(1):83-92. [doi: 10.1097/01.AOG.0000218100.05601.f8] [Medline: 16816060]
174. Peterson AV, Kealey KA, Mann SL, Marek PM, Ludman EJ, Liu J, et al. Group-randomized trial of a proactive, personalized telephone counseling intervention for adolescent smoking cessation. J Natl Cancer Inst 2009 Oct 21;101(20):1378-1392 [FREE Full text] [doi: 10.1093/jnci/djp317] [Medline: 19822836]
175. Severson HH, Peterson AL, Andrews JA, Gordon JS, Cigrang JA, Danaher BG, et al. Smokeless tobacco cessation in military personnel: a randomized controlled trial. Nicotine Tob Res 2009 Jun;11(6):730-738. [doi: 10.1093/ntr/ntp057] [Medline: 19395686]
176. Bastian LA, Fish LJ, Peterson BL, Biddle AK, Garst J, Lyna P, et al. Assessment of the impact of adjunctive proactive telephone counseling to promote smoking cessation among lung cancer patients' social networks. Am J Health Promot 2013;27(3):181-190. [doi: 10.4278/ajhp.101122-QUAN-387] [Medline: 23286595]
177. Jiménez-Muro A, Nerín I, Samper P, Marqueta A, Beamonte A, Gargallo P, et al. A proactive smoking cessation intervention in postpartum women. Midwifery 2013 Mar;29(3):240-245. [doi: 10.1016/j.midw.2012.01.003] [Medline: 22361008]
178. Woodruff SI, Conway TL, Edwards CC, Elliott SP, Crittenden J. Evaluation of an Internet virtual world chat room for adolescent smoking cessation. Addict Behav 2007 Sep;32(9):1769-1786. [doi: 10.1016/j.addbeh.2006.12.008] [Medline: 17250972]
179. Becker J, Haug S, Sullivan R, Schaub MP. Effectiveness of different Web-based interventions to prepare co-smokers of cigarettes and cannabis for double cessation: a three-arm randomized controlled trial. J Med Internet Res 2014 Dec 05;16(12):e273 [FREE Full text] [doi: 10.2196/jmir.3246] [Medline: 25486674]
180. Mason MJ, Campbell L, Way T, Keyser-Marcus L, Benotsch E, Mennis J, et al. Development and outcomes of a text messaging tobacco cessation intervention with urban adolescents. Subst Abus 2015;36(4):500-506. [doi: 10.1080/08897077.2014.987946] [Medline: 25551337]
181. Blankers M, Koeter MW, Schippers GM. Internet therapy versus internet self-help versus no treatment for problematic alcohol use: a randomized controlled trial. J Consult Clin Psychol 2011 Jun;79(3):330-341 [FREE Full text] [doi: 10.1037/a0023498] [Medline: 21534652]
182. Brown RL, Saunders LA, Bobula JA, Mundt MP, Koch PE. Randomized-controlled trial of a telephone and mail intervention for alcohol use disorders: three-month drinking outcomes. Alcohol Clin Exp Res 2007 Aug;31(8):1372-1379. [doi: 10.1111/j.1530-0277.2007.00430.x] [Medline: 17550366]
183. Borsari B, Short EE, Mastroleo NR, Hustad JTP, Tevyaw TO, Barnett NP, et al. Phone-delivered brief motivational interventions for mandated college students delivered during the summer months. J Subst Abuse Treat 2014 May;46(5):592-596 [FREE Full text] [doi: 10.1016/j.jsat.2014.01.001] [Medline: 24512944]

184. LaChance H, Feldstein ES, Bryan AD, Hutchison KE. What makes group MET work? A randomized controlled trial of college student drinkers in mandated alcohol diversion. Psychol Addict Behav 2009 Dec;23(4):598-612 [FREE Full text] [doi: 10.1037/a0016633] [Medline: 20025366]
185. Wongpakaran T, Petcharaj K, Wongpakaran N, Sombatmai S, Boripuntakul T, Intarakamhaeng D, et al. The effect of telephone-based intervention (TBI) in alcohol abusers: a pilot study. J Med Assoc Thai 2011 Jul;94(7):849-856. [Medline: 21774293]
186. D'Amico EJ, Hunter SB, Miles JNV, Ewing BA, Osilla KC. A randomized controlled trial of a group motivational interviewing intervention for adolescents with a first time alcohol or drug offense. J Subst Abuse Treat 2013;45(5):400-408 [FREE Full text] [doi: 10.1016/j.jsat.2013.06.005] [Medline: 23891459]
187. D'Amico EJ, Houck JM, Hunter SB, Miles JNV, Osilla KC, Ewing BA. Group motivational interviewing for adolescents: change talk and alcohol and marijuana outcomes. J Consult Clin Psychol 2015 Feb;83(1):68-80 [FREE Full text] [doi: 10.1037/a0038155] [Medline: 25365779]
188. Nyamathi A, Shoptaw S, Cohen A, Greengold B, Nyamathi K, Marfisee M, et al. Effect of motivational interviewing on reduction of alcohol use. Drug Alcohol Depend 2010 Feb 01;107(1):23-30 [FREE Full text] [doi: 10.1016/j.drugalcdep.2009.08.021] [Medline: 19836904]
189. Nyamathi AM, Nandy K, Greengold B, Marfisee M, Khalilifard F, Cohen A, et al. Effectiveness of intervention on improvement of drug use among methadone maintained adults. J Addict Dis 2011 Jan;30(1):6-16 [FREE Full text] [doi: 10.1080/10550887.2010.531669] [Medline: 21218306]
190. Suffoletto B, Kristan J, Callaway C, Kim KH, Chung T, Monti PM, et al. A text message alcohol intervention for young adult emergency department patients: a randomized clinical trial. Ann Emerg Med 2014 Dec;64(6):664-72.e4 [FREE Full text] [doi: 10.1016/j.annemergmed.2014.06.010] [Medline: 25017822]
191. Suffoletto B, Kristan J, Chung T, Jeong K, Fabio A, Monti P, et al. An interactive text message intervention to reduce binge drinking in young adults: a randomized controlled trial with 9-Month Outcomes. PLoS One 2015;10(11):e0142877 [FREE Full text] [doi: 10.1371/journal.pone.0142877] [Medline: 26580802]
192. Gates PJ, Norberg MM, Copeland J, Digiusto E. Randomized controlled trial of a novel cannabis use intervention delivered by telephone. Addiction 2012 Dec;107(12):2149-2158. [doi: 10.1111/j.1360-0443.2012.03953.x] [Medline: 22632139]
193. Madigan K, Brennan D, Lawlor E, Turner N, Kinsella A, O'Connor JJ, et al. A multi-center, randomized controlled trial of a group psychological intervention for psychosis with comorbid cannabis dependence over the early course of illness. Schizophr Res 2013 Jan;143(1):138-142. [doi: 10.1016/j.schres.2012.10.018] [Medline: 23187069]
194. Stevens J, Hayes J, Pakalnis A. A randomized trial of telephone-based motivational interviewing for adolescent chronic headache with medication overuse. Cephalalgia 2014 May;34(6):446-454. [doi: 10.1177/0333102413515336] [Medline: 24322483]
195. Agyapong VIO, Ahern S, McLoughlin DM, Farren CK. Supportive text messaging for depression and comorbid alcohol use disorder: single-blind randomised trial. J Affect Disord 2012 Dec 10;141(2-3):168-176. [doi: 10.1016/j.jad.2012.02.040] [Medline: 22464008]
196. Weitzel JA, Bernhardt JM, Usdan S, Mays D, Glanz K. Using wireless handheld computers and tailored text messaging to reduce negative consequences of drinking alcohol. J Stud Alcohol Drugs 2007 Jul;68(4):534-537. [Medline: 17568957]
197. Suffoletto B, Callaway C, Kristan J, Kraemer K, Clark DB. Text-message-based drinking assessments and brief interventions for young adults discharged from the emergency department. Alcohol Clin Exp Res 2012 Mar;36(3):552-560. [doi: 10.1111/j.1530-0277.2011.01646.x] [Medline: 22168137]
198. Haug S, Schaub MP, Venzin V, Meyer C, John U, Gmel G. A pre-post study on the appropriateness and effectiveness of a Web- and text messaging-based intervention to reduce problem drinking in emerging adults. J Med Internet Res 2013;15(9):e196 [FREE Full text] [doi: 10.2196/jmir.2755] [Medline: 23999406]
199. Gajecki M, Berman AH, Sinadinovic K, Rosendahl I, Andersson C. Mobile phone brief intervention applications for risky alcohol use among university students: a randomized controlled study. Addict Sci Clin Pract 2014;9:11 [FREE Full text] [doi: 10.1186/1940-0640-9-11] [Medline: 24985342]
200. Gonzales R, Ang A, Murphy DA, Glik DC, Anglin MD. Substance use recovery outcomes among a cohort of youth participating in a mobile-based texting aftercare pilot program. J Subst Abuse Treat 2014 Jul;47(1):20-26 [FREE Full text] [doi: 10.1016/j.jsat.2014.01.010] [Medline: 24629885]
201. 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. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
202. Lucht MJ, Hoffman L, Haug S, Meyer C, Pussehl D, Quellmalz A, et al. A surveillance tool using mobile phone short message service to reduce alcohol consumption among alcohol-dependent patients. Alcohol Clin Exp Res 2014 Jun;38(6):1728-1736. [doi: 10.1111/acer.12403] [Medline: 24730528]
203. Shrier LA, Rhoads A, Burke P, Walls C, Blood EA. Real-time, contextual intervention using mobile technology to reduce marijuana use among youth: a pilot study. Addict Behav 2014 Jan;39(1):173-180. [doi: 10.1016/j.addbeh.2013.09.028] [Medline: 24139665]

204. Haug S, Lucht MJ, John U, Meyer C, Schaub MP. A pilot study on the feasibility and acceptability of a text message-based aftercare treatment programme among alcohol outpatients. Alcohol Alcohol 2015 Mar;50(2):188-194. [doi: 10.1093/alcalc/agu107] [Medline: 25600249]
205. Gonzalez VM, Dulin PL. Comparison of a smartphone app for alcohol use disorders with an internet-based intervention plus bibliotherapy: A pilot study. J Consult Clin Psychol 2015 Apr;83(2):335-345. [doi: 10.1037/a0038620] [Medline: 25622202]
206. Schwinn TM, Schinke SP, Di Noia J. Preventing drug abuse among adolescent girls: outcome data from an internet-based intervention. Prev Sci 2010 Mar;11(1):24-32 [FREE Full text] [doi: 10.1007/s11121-009-0146-9] [Medline: 19728091]
207. Wood S, Eckley L, Hughes K, Hardcastle K, Bellis M, Schrooten J, et al. Computer-based programmes for the prevention and management of illicit recreational drug use: a systematic review. Addict Behav 2014 Jan;39(1):30-38. [doi: 10.1016/j.addbeh.2013.09.010] [Medline: 24144590]
208. Hopson L, Wodarski J, Tang N. The effectiveness of electronic approaches to substance abuse prevention for adolescents. J Evid Inf Soc Work 2015;12(3):310-322. [Medline: 25661894]
209. Champion KE, Newton NC, Barrett EL, Teesson M. A systematic review of school-based alcohol and other drug prevention programs facilitated by computers or the internet. Drug Alcohol Rev 2013 Mar;32(2):115-123. [doi: 10.1111/j.1465-3362.2012.00517.x] [Medline: 23039085]
210. Newton NC, Andrews G, Champion KE, Teesson M. Universal Internet-based prevention for alcohol and cannabis use reduces truancy, psychological distress and moral disengagement: a cluster randomised controlled trial. Prev Med 2014 Aug;65:109-115. [doi: 10.1016/j.ypmed.2014.05.003] [Medline: 24823906]
211. Champion KE, Newton NC, Stapinski L, Slade T, Barrett EL, Teesson M. A cross-validation trial of an Internet-based prevention program for alcohol and cannabis: preliminary results from a cluster randomised controlled trial. Aust N Z J Psychiatry 2016 Jan;50(1):64-73. [doi: 10.1177/0004867415577435] [Medline: 25801662]
212. Champion KE, Newton NC, Stapinski LA, Teesson M. Effectiveness of a universal internet-based prevention program for ecstasy and new psychoactive substances: a cluster randomized controlled trial. Addiction 2016 Apr 13;111(8):1396-1405. [doi: 10.1111/add.13345]
213. Christoff AO, Boerngen-Lacerda R. Reducing substance involvement in college students: a three-arm parallel-group randomized controlled trial of a computer-based intervention. Addict Behav 2015 Jun;45:164-171. [doi: 10.1016/j.addbeh.2015.01.019] [Medline: 25679364]
214. Champion K, Newton N, Teesson M. Prevention of alcohol and other drug use and related harm in the digital age: what does the evidence tell us? Curr Opin Psychiatry 2016 Dec;29(4):242-249. [doi: 10.1097/YCO.0000000000000258] [Medline: 27153124]
215. Brinn MP, Carson KV, Esterman AJ, Chang AB, Smith BJ. Mass media interventions for preventing smoking in young people. Cochrane Database Syst Rev 2010 Nov 10(11):CD001006. [doi: 10.1002/14651858.CD001006.pub2] [Medline: 21069667]
216. Carson KV, Ameer F, Sayehmiri K, Hnin K, van Agteren JE, Sayehmiri F, et al. Mass media interventions for preventing smoking in young people. Cochrane Database Syst Rev 2017 Dec 02;6:CD001006. [doi: 10.1002/14651858.CD001006.pub3] [Medline: 28574573]
217. Eiser C, Eiser J, Pritchard M. Reactions to Drug Education: a comparison of two videos produced for schools. Br J Addict 1988 Aug;83(8):955-963. [doi: 10.1111/j.1360-0443.1988.tb01589.x]
218. Haug S, Paz Castro R, Meyer C, Filler A, Kowatsch T, Schaub M. A mobile phone-based life skills training program for substance use prevention among adolescents: pre-post study on the acceptance and potential effectiveness of the program, Ready4life. JMIR Mhealth Uhealth 2017 Oct 04;5(10):e143 [FREE Full text] [doi: 10.2196/mhealth.8474] [Medline: 28978498]
219. Deitz D, Cook R, Hendrickson A. Preventing prescription drug misuse: field test of the SmartRx Web program. Subst Use Misuse 2011;46(5):678-686. [doi: 10.3109/10826084.2010.528124] [Medline: 21043788]
220. Fang L, Schinke S, Cole K. Preventing substance use among early Asian-American adolescent girls: initial evaluation of a web-based, mother-daughter program. J Adolesc Health 2010 Nov;47(5):529-532 [FREE Full text] [doi: 10.1016/j.jadohealth.2010.03.011] [Medline: 20970090]
221. Terry-McElrath YM, Emery S, Szczypka G, Johnston LD. Potential exposure to anti-drug advertising and drug-related attitudes, beliefs, and behaviors among United States youth, 1995-2006. Addict Behav 2011;36(1-2):116-124 [FREE Full text] [doi: 10.1016/j.addbeh.2010.09.005] [Medline: 20961691]
222. Brannon BR, Dent CW, Flay BR, Smith G, Sussman S, Pentz MA, et al. The television, school, and family project. V. The impact of curriculum delivery format on program acceptance. Prev Med 1989 Jul;18(4):492-502. [Medline: 2798371]
223. Schinke S, Fang L, Cole K, Cohen-Cutler S. Preventing substance use among Black and Hispanic adolescent girls: results from a computer-delivered, mother-daughter intervention approach. Subst Use Misuse 2011;46(1):35-45 [FREE Full text] [doi: 10.3109/10826084.2011.521074] [Medline: 21190404]
224. Klisch Y, Bowling KG, Miller LM, Ramos MA. The impact of science education games on prescription drug abuse attitudes among teens: a case study. J Drug Educ 2013 Aug;43(3):255-275. [doi: 10.2190/DE.43.3.d] [Medline: 25445507]

225. Schuman SH. Drug perception and the student-teacher gap. Reactions of 428 students and 72 teachers to an experimental trigger film on drugs. J Am Med Assoc 1971 Apr 26;216(4):659-663. [Medline: 4101684]
226. Flay BR. Mass media linkages with school-based programs for drug abuse prevention. J Sch Health 1986 Nov;56(9):402-406. [Medline: 3537525]
227. Schinke SP, Fang L, Cole KC. Computer-delivered, parent-involvement intervention to prevent substance use among adolescent girls. Prev Med 2009 Nov;49(5):429-435 [FREE Full text] [doi: 10.1016/j.ypmed.2009.08.001] [Medline: 19682490]
228. 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] [Medline: 19067165]
229. Andrews J, Gordon J, Hampson S, Gunn B, Christiansen S, Slovic P. Long-term efficacy of click city(r): tobacco: a school-based tobacco prevention program. Nicotine Tob Res 2014 Jan;16(1):33-41 [FREE Full text] [doi: 10.1093/ntr/ntt106] [Medline: 23884322]
230. Kinder BN. Attitudes toward alcohol and drug abuse. II. Experimental data, mass media research, and methodological considerations. International Journal of the Addictions 2009 Jul 03;10(6):1035-1054. [doi: 10.3109/10826087509028359]
231. Bandy P, President P. Recent literature on drug abuse prevention and mass media: focusing on youth, parents, women and the elderly. J Drug Educ 1995 Jan;13(3):255-271. [doi: 10.2190/2GXA-K970-5L68-RE14]
232. Johnson E, Delgado J. Reaching Hispanics with messages to prevent alcohol and other drug abuse. Public Health Rep 1989;104(6):588-594 [FREE Full text] [Medline: 2479957]
233. Jiang S, Wu L, Gao X. Beyond face-to-face individual counseling: a systematic review on alternative modes of motivational interviewing in substance abuse treatment and prevention. Addict Behav 2017 Dec;73:216-235. [doi: 10.1016/j.addbeh.2017.05.023] [Medline: 28554033]
234. Schinke SP, Fang L, Cole KC. Preventing substance use among adolescent girls: 1-year outcomes of a computerized, mother-daughter program. Addict Behav 2009 Dec;34(12):1060-1064 [FREE Full text] [doi: 10.1016/j.addbeh.2009.06.007] [Medline: 19632053]
235. Schwinn T, Schinke S, Hopkins J, Keller B, Liu X. An online drug abuse prevention program for adolescent girls: posttest and 1-year outcomes. J Youth Adolesc 2018 Dec;47(3):490-500. [doi: 10.1007/s10964-017-0714-4] [Medline: 28755247]
236. Moncher MS, Parms CA, Orlandi MA, Schinke SP, Miller SO, Palleja J, et al. Microcomputer-based approaches for preventing drug and alcohol abuse among adolescents from ethnic-racial minority backgrounds. Comput Human Behav 1989 Jan;5(2):79-93. [doi: 10.1016/0747-5632(89)90026-5]
237. Flay BR, Sobel JL. The role of mass media in preventing adolescent substance abuse. NIDA Res Monogr 1983;47:5-35. [Medline: 6419119]
238. Wallack L. Mass media campaigns: the odds against finding behavior change. Health Educ Q 2016 Sep 04;8(3):209-260. [doi: 10.1177/109019818100800302]
239. Wallack L. Mass media and drinking, smoking, and drug taking. In: Contemporary drug problems. United States: Elsevier; 1980:49-83.
240. Barcus F, Jankowski S. Drugs and the mass media. Ann Am Acad Polit Soc Sci 2016 Sep 08;417(1):86-100. [doi: 10.1177/000271627541700109]
241. Andrews J, Gordon J, Hampson S, Christiansen S, Gunn B, Slovic P, et al. Short-term efficacy of Click City®: Tobacco: changing etiological mechanisms related to the onset of tobacco use. Prev Sci 2011 Mar;12(1):89-102 [FREE Full text] [doi: 10.1007/s11121-010-0192-3] [Medline: 21286810]
242. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432. [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
243. The Evolution of Media. In: Understanding Media and Culture: An Introduction to Mass Communication. United States: University of Minnesota Libraries Publishing edition; 2016:22-21.
244. Marsch L, Bickel W, Badger G. Applying computer technology to substance abuse prevention science: results of a preliminary examination. J Child Adolesc Subst Abuse 2007 Mar 06;16(2):69-94. [doi: 10.1300/J029v16n02_04]
245. U.S. Department of Health and Human Services. HHS.GOV/OPIOIDS. 2018. What is the U.S. Opioid Epidemic? URL: https://www.hhs.gov/opioids/about-the-epidemic/index.html [accessed 2018-07-20] [WebCite Cache ID 713YWM9Mh]

Abbreviations

CHIT: consumer health information technology
mHealth: mobile health
NPS: new psychoactive substances
RCT: randomized controlled trials
STDs: sexually transmitted diseases
WHO: World Health Organization

Edited by G Eysenbach; submitted 03.08.18; peer-reviewed by K Champion, S Tao, S Kershaw; comments to author 20.09.18; revised version received 08.10.18; accepted 20.10.18; published 30.01.19.

Please cite as:
Pradhan AM, Park L, Shaya FT, Finkelstein J
Consumer Health Information Technology in the Prevention of Substance Abuse: Scoping Review
J Med Internet Res 2019;21(1):e11297
URL: http://www.jmir.org/2019/1/e11297/
doi:10.2196/11297
PMID:30698526

Viewpoint

E-Learning for Medical Education in Sub-Saharan Africa and Low-Resource Settings: Viewpoint

Sandra Barteit[1], MA; Albrecht Jahn[1], MD, PhD, MSc; Sekelani S Banda[2], MB ChB, MSc, MMed, PhD; Till Bärnighausen[1,3,4], MD, PhD, MSc, ScD; Annel Bowa[5], MSc; Geoffrey Chileshe[5], BSc; Dorota Guzek[1], MSc; Margarida Mendes Jorge[1], MSc; Sigrid Lüders[6], MD; Gregory Malunga[5], MSc; Florian Neuhann[1], MD

[1]Heidelberg Institute of Global Health, Heidelberg, Germany

[2]Ministry of Health, Lusaka, Zambia

[3]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[4]Africa Health Research Institute (AHRI), KwaZulu-Natal, South Africa

[5]Chainama College of Health Sciences, Lusaka, Zambia

[6]SolidarMed, Lusaka, Zambia

Corresponding Author:
Sandra Barteit, MA
Heidelberg Institute of Global Health
Im Neuenheimer Feld 130.3
Heidelberg, 69120
Germany
Phone: 49 06221 ext 5634030
Email: barteit@uni-heidelberg.de

Abstract

E-learning has been heralded as a revolutionary force for medical education, especially for low-resource countries still suffering from a dire lack of health care workers. However, despite over two decades of e-learning endeavors and interventions across sub-Saharan Africa and other low- and middle-income countries, e-learning for medical education has not gained momentum and continues to fall short of the anticipated revolution. Many e-learning interventions have been cul-de-sac pilots that have not been scaled up but rather terminated after the pilot phase. This is usually a result of not adopting a system-wide approach, which leads to insufficient scope of training, insufficient technological maintenance and user support, unattainably high expectations, and unrealistic financial planning. Thus, a multitude of e-learning evaluations have failed to provide scientifically sound evidence of the effectiveness of e-learning for medical education in low-resource countries. Instead, it appears that technological development has overwhelmed rather than revolutionized medical education. The question of how to push e-learning into a higher gear in low-resource countries persists. Provision of e-learning as a technology is insufficient. E-learning needs to be vigorously and sustainably integrated into the local educational setting and aligned with national strategies and other national endeavors and interventions. Adhering to a standardized framework for the implementation and evaluation of e-learning endeavors is key, especially to bridge the gap in robust evidence that should also guide e-learning implementations. The primary objective of e-learning for medical education is to strengthen the health system in order to serve the population's health care needs and expectations. Currently, medical e-learning does not measure up to its potential or do justice to medical students in low-resource countries. Technology may help unfold the potential of e-learning, but an all-encompassing change is needed. This can only be achieved through a joint effort that follows a systematic and standardized framework, especially for implementation and evaluation.

(J Med Internet Res 2019;21(1):e12449) doi:10.2196/12449

KEYWORDS

medical e-learning; technology-enhanced learning; blended learning; health workers; health system strengthening; universal health coverage; medical education; mHealth; eHealth; developing countries; sub-Saharan Africa; low-resource countries

Introduction

> *The effect of electronic learning (e-learning) is likely to be revolutionary, although how precisely it will revamp professional education remains unknown.*

This statement is part of a vision and strategy by Frenk et al (2010) for a commission on education and health workers for the 21st century [1]. As of 2018, the e-learning revolution in low-resource countries has not yet taken off, and the potential of e-learning in supporting the advancement of health training is still unknown.

What are the barriers that currently restrict the potential of e-learning for medical training in low-resource countries? How can e-learning address the continuing crisis in human resources for health? It is clear that progress is profoundly needed, considering that low-income countries are faced with fragile health systems and dramatically insufficient numbers of health workers. Of the 49 nations in sub-Saharan Africa with approximately one billion people, only about 6,000 medical doctors graduate per year as compared to Western Europe with a population of 200 million (one-fifth of sub-Saharan Africa), where the number of doctors that graduate per year (42,000) is seven times that in sub-Saharan Africa [1]. Low graduation numbers in sub-Saharan Africa reflect the small number of medical schools (134) in comparison to Western Europe (almost 300) [1]. As such, sub-Saharan Africa has an urgent need to scale up existing educational infrastructure to meet the large and increasing number of young people striving for education and to address the dire need for healthcare in their population by educating higher numbers of qualified health personnel. The current educational infrastructure (in particular, medical education), practical training for health workers, and continuous medical education are inadequate in quantity and quality. Institutions for medical education such as colleges of nursing, allied health sciences, and medical schools are faced with a limited capacity to enroll students and inadequate numbers of medical educators and are also affected by a lack of infrastructure such as an insufficient number of classrooms, lecture halls, or dorms.

Technology, on the other hand, is developing by leaps and bounds. Personal computers were introduced for the general public in the 1970s, and throughout the recent decades, technological advancement has been astonishingly rapid. Today, computers are widely available in many forms: tablets, mobile phones, laptops, and virtual reality headsets. These technologies change educational approaches and can potentially improve health care delivery. However, global digital resources are unequally distributed, causing a so-called “digital divide.” To empower low-income countries to fast-track the lengthy early development stages that high-income countries experience [1], narrowing of the digital gap by increasing the development of and access to technological resources is required. One fast-track technological advance is exemplified by the rapid and extensive coverage of mobile phones in sub-Saharan Africa [2]. Overall, sub-Saharan countries have progressed significantly in the past few years in mobile connectivity and internet access. This progress provides a fertile setting for medical e-learning; its advantages include flexible learning, time efficiency, potential lower costs (due to a reduced need for printed learning materials such as books, and easily updated e-materials), standardization of course content, distance delivery, and scalability. E-learning for medical education might provide these countries a chance to increase the numbers of trained health workers in low-income countries while maintaining or potentially improving the quality of education with the support of self-directed learning and thus decrease the workload of current health workers engaged in health education.

Pilotitis: How to Overcome Pilots

Published literature has documented a multitude of e-learning interventions for medical education and electronic health (eHealth) in sub-Saharan Africa. In fact, there have been so many eHealth projects, that in 2012, Uganda issued a directive to halt all national mobile health (mHealth) initiatives and stopped pilots that used mobile and wireless devices for health across the country [3] despite the heralded revolution of health technology and its promised potential. Uganda reacted to the perceived chaos surrounding the pilots and paused to develop standards for technical specifications and align projects with Uganda’s national health strategy. Uganda is indicative of the general trend in sub-Saharan Africa in terms of e-learning for medical education using a myriad of approaches. E-learning interventions tend to be isolated or running in parallel in a single country, which evokes an impression of urgency, resulting in imprudent implementation based on a strategy of “anything is better than nothing.” The recent Global Observatory for eHealth of the World Health Organization [4] found that the overall number of pilots are decreasing and the number of implemented large-scaled projects are increasing; however, the recent report of the Broadband Commission for Sustainable Development states that there is still no national coordination of digital health solutions in low- and middle-income countries, leading to a “fragmented ecosystem” [5]. In addition, many e-learning interventions do not take a system-wide approach [6], leaving out important aspects in the e-learning design plan. At a later stage, such exclusions evolve to major barriers of technology adoption and lead to frustrations and failure of implementing the intervention beyond the pilot phase, such as insufficient training of all stakeholders (students, teachers, and administration), insufficient maintenance and technological user support, unattainably high expectations, and unrealistic financial planning [7]. As a result, many of these technology-enhanced interventions and projects never progress past the pilot stage [5]; they end before scale-up—a phenomenon named *pilotitis* [6]. Clearly, pilots are important as a small-scale testing phase before investing in a large-scale deployment, but the pilot requires clear-defined objectives paired with monitoring and evaluation [7]. Sustainable implementation leading to scale-up is crucial for medical e-learning to realize its full potential.

Structured and Sustainable Implementation

To be successful, e-learning interventions need to follow a systemic approach in a given educational environment and more

specifically consider the definition of scope, objectives and target group, availability of a corresponding curriculum, active involvement of teachers and administrators, sufficient information technology (IT) support, adequate IT infrastructure, and clear political and institutional support.

A standard evaluation framework can provide insight into adapting medical e-learning pilots to local conditions and needs. Interventions should be interwoven as tightly as possible within the local educational infrastructure [8]. At the infrastructural core are individuals who are committed learners; teachers; and administrative and support staff such as the head of the department, teaching coordinators, IT support, and network administrators. The curriculum provides guidance at the individual level, which is crucial for e-learning interventions for medical education, especially for integration of e-learning into the curriculum. However, as the commission on education and health professionals for the 21st century has noted, "professional education has not kept pace...largely because of fragmented, outdated, and static curricula that produce ill-equipped graduates," as "local educational standards are all too often driven by the desire to fit into frameworks that are in place elsewhere" [1]. Thus, the introduction of e-learning can be a chance to overhaul current curricula to consider and integrate scientific advancements, learner-centered models that are competence oriented, and technology-enhanced teaching and learning methods.

The internet and other technologies have fostered the growth of knowledge that is generally available free of cost for anyone with the technological means of access. Thus, nowadays, knowledge is accessible like never before. Just a few decades ago, when most current medical curricula were developed, memorization of facts was often considered paramount. Today, the skill to locate necessary information has become more critical as part of synthesis, analysis, and decision-making processes [1]. With the advent of new learning technologies and changes in knowledge handling, educational institutions in low-income countries should use this opportunity to update their curricula [1]. The incorporation of e-learning in the curriculum would confirm its role as a significant educational device to foster progress, rather than the current perception of e-learning as a "technological toy." In low-income countries facing insufficient numbers of medical teachers and in need of making fast improvements, e-learning could be an essential step to relieve the capacity overload of the small number of medical teachers by embracing new technological approaches in order to acquire knowledge and skills. For example, instead of a teacher-centered approach, a blended learning approach could be put in place wherein students learn with self-directed e-learning materials such as interactive quizzes, videos, and literature to deepen the understanding and cover a range of topics or modules. This e-learning component could be blended with face-to-face exchanges and thus decrease the time invested by medical teachers by substituting study discussion groups or a flipped classroom, wherein the medical teacher facilitates a discussion and a question-and-answer session. This would remove the need and time invested to prepare full lectures, but still target students' needs with question-and-answer sessions.

The curriculum change is a crucial element for successful e-learning integration, but it is only one piece of the "puzzle." The successful implementation and sound evaluation must follow a multilevel approach that incorporates the individual learner, the learning environment, the context of the e-learning implementation, the technological environment, and the pedagogics involved in the e-learning implementation [9].

Before starting an e-learning implementation, the objectives and expected outcomes of the intervention should be clearly defined. On an individual level, a needs assessment can provide valuable insights for implementation regarding the content, e-learning design, and technological equipment. This would answer questions about what content the learners expect and what content can be offered. Materials should be targeted at the learner, consider cultural context, and align with national health strategies and guidelines. To develop learning materials, the input and active involvement of medical staff is needed, which constitute a bottleneck because e-learning often aims to bridge the lack of medical teachers. Hence, a strong commitment to a substantial initial investment in medical e-learning is required from medical teachers, institutional administration, and governing bodies [8]. Without convincing and contextual contents, medical e-learning remains a skeleton without meaningful outcomes.

Another important aspect is how the medical e-learning content should be designed and how learners and teachers can best use the e-learning materials. For content design, the users' needs should be evaluated and a pedagogical e-learning strategy according to the curriculum should be established. Current learning-management systems such as the popular and open-source software Moodle [10] already offer a pedagogical framework that can be assembled to local needs and standards. The content should follow a standard model such as SCORM (shareable content object reference model) [11] to ensure compatibility with other software platforms and enable potential sharing of content. Pros and cons of what technology is best to employ should be carefully balanced within the given setting, especially with regard to the users (learners and teachers), support services (information and communication technology staff), and technological sustainability including long-term maintenance options and technological utility in the given low-resource setting. Are the users able to use the device without too much effort? Can the technological device be repaired in cases of damage within the country? Is the technological device fit for the environment? It may be necessary to provide technical devices to learners and teachers to ensure equal access to medical e-learning. Is access mainly online or offline? Should the e-learning platform be provided via a local server or a rented service? Who can maintain and update the technological infrastructure? The answers to these questions should guide the selection of technical devices and services. Financial planning also needs to consider reoccurring costs for technical equipment, ICT training domain-name registration, and data safety and confidentiality.

One remedy against *pilotitis* could be the technology itself. For example, an online database similar to a clinical trial registry could be established, by which eHealth and e-learning interventions for sub-Saharan Africa (or even globally) are

centrally registered. This database could provide transparency to the current black box of eHealth and medical e-learning interventions. Potentially, existing institutions and efforts, for example, the Broadband Commission for Sustainable Development or the World Health Organization as part of the Global Observatory for eHealth, could provide frameworks and potentially, resources to host, maintain, and develop such an eHealth and medical e-learning database. This would ease collaboration, planning, and priority setting on national and international levels. The most-effective interventions among ongoing interventions could then be implemented where they are needed the most. Parameters such as the type of technology used for the intervention as well as evaluation methods and measured outcomes could be integrated into the database, thus supporting the advancement of standard evaluation methods for medical e-learning. This database could be a point of entry for new stakeholders as well as inform policymakers and regulatory bodies. National policies should support medical e-learning interventions and provide the necessary framework for structured growth of e-learning initiatives.

The Alpha and Omega of E-Learning: Evaluation

Many published medical e-learning evaluations have low-quality scientific standards [12] and often report their results in a narrative manner [12-14] without following a reporting standard for either qualitative or quantitative evaluation. The majority of e-learning interventions rely on self-designed evaluation designs that are rarely validated. Subjective evidence generally focuses solely on the individual learners, such as learner satisfaction or other user-perceived parameters (eg, learner opinion). Studies that only incorporate the individual learner in their evaluation are restricted in their insight into the intervention outcome. Consequently, a tailored adaptation to the learners' actual needs and requirements is difficult to achieve. Another limitation of this approach is that evidence from published medical e-learning studies cannot be used for comparison to other e-learning studies [12-14] and lessons that may have been learned from comparison are lost. Publications have tackled e-learning evaluation concepts and models [15-17] by emphasizing the need to comprehensively evaluate all levels of an e-learning intervention—from the learner to the institutional and governmental contexts. However, the heterogeneity of studies and the prominence of subjective evidence has led to a continued lack of sound scientific evidence on the effect of e-learning in medical education in low-resource countries. Therefore, a comprehensive evaluation of medical e-learning interventions is needed to generate causal evidence. The reasons for the predominance of trial and error pilots and the prevalence of "homemade" e-learning evaluations remain unclear. Resources to guide implementation and evaluation of e-learning interventions are plenty. Although fully randomized controlled trials might be difficult to conduct in regulated systems such as medical education, other causal research designs can and should be employed. Many methods like the objective structured clinical examinations or standardized multiple-choice tests are available for skill and knowledge testing. Prevalent models in e-learning for questionnaires include the technology acceptance models (TAM1 and TAM2) for assessing technology acceptance, the unified theory of acceptance and use, and the system usability scale for usability.

The CONSORT (Consolidated Standards of Reporting Trials) statement and NOS-E (Newcastle-Ottawa Scale for Education) include standards for reporting and assessing study quality. The primary objective for most interventions is to employ e-learning to strengthen medical education and provide good-quality education. Therefore, the implementation and evolution of e-learning should target these objectives and produce evidence that allows for causal conclusions.

Future Work

In our opinion, a fundamental requirement of e-learning for medical education should be systematic, especially with regard to planning, implementation, and evaluation. Adherence to a general and standard implementation and evaluation framework that can be further refined for the local setting and employment of more causal research designs could improve sustainability and the quality of evidence. However, e-learning for medical education in low-resource countries still requires evidence to prove equivalence or even superiority to analogue and traditional educational approaches. A clear strategy for adapted implementation and integration as well as the understanding that e-learning is not a cure for all, but rather a valuable tool requiring a strong commitment and significant upfront investment with continued support, is necessary. In particular, a definition of new roles and potentially, the creation of new staff positions are needed. To move forward with this new structure, political will is essential for establishing national and regional strategies. Thus, international partnerships may provide a strong foundation of resources for initial setup. Alongside the digital educational infrastructure with e-learning for medical education in low-resource countries, health systems can be structurally and substantially strengthened with eHealth initiatives.

Conclusions

E-learning for medical education in low-resource countries is in a disorganized state, with many pilot projects that are not scaled up. Therefore, the potential of e-learning, especially for medical education, remains underutilized. Making technology available is not sufficient nor does it do justice to the many potential general and medical students who need quality education in sub-Saharan Africa. We are not suggesting that e-learning is the magic bullet to solve existing problems with medical training in resource-limited settings with a lack of health workers and qualified medical educators. However, we see e-learning as an important and potent component with a potential that remains mostly underexploited to date in these settings. Therefore, we propose a concerted effort to unfold and enhance the effectiveness of e-learning as an educational tool to increase the quantity and quality of medical education programs. We suggest (1) a topic-specific database that registers all e-learning and eHealth interventions similar to clinical trial registers, (2) a standardized and widely employed framework for the evaluation of e-learning programs, and (3) structured programs

that incorporate e-learning within and between medical teaching institutions and accreditation bodies. E-learning for medical education is not a self-runner: It requires significant upfront investment in content, training, and technology as well as the acknowledgment of recurring costs that will potentially be paid off only at a later stage [8]. For low-resource countries struggling to increase the number of health workers, medical e-learning may be able to accelerate progress by skipping many previous developmental steps taken by high-income countries in the past. Available technology, including software and hardware, has the potential to fundamentally change and shape health systems, the quality of health education, and subsequently, the quality and quantity of health care delivery and access. Stakeholders need to focus on formulating and promoting standard frameworks for implementation and evaluation [16]. Health is about people and so is medical e-learning. We need to emphasize that education is at the core of strengthening health systems as education enhances the performance of health systems, so that they can meet the needs of patients and populations equitably and efficiently.

Acknowledgments

We thank Julia Challinor for her editorial support.

Conflicts of Interest

None declared.

Authors' Contributions

SB wrote the full draft of the manuscript. FN revised and reviewed the initial draft. All authors read, edited, and approved the final manuscript.

References

1. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. The Lancet 2010 Dec 04;376(9756):1923-1958. [doi: 10.1016/S0140-6736(10)61854-5] [Medline: 21112623]
2. Hampshire K, Porter G, Owusu S, Mariwah S, Abane A, Robson E, et al. Informal m-health: How are young people using mobile phones to bridge healthcare gaps in Sub-Saharan Africa? Soc Sci Med 2015 Oct;142:90-99 [FREE Full text] [doi: 10.1016/j.socscimed.2015.07.033] [Medline: 26298645]
3. Franz-Vasdeki J, Pratt B, Newsome M, Germann S. Taking mHealth Solutions to Scale: Enabling Environments and Successful Implementation. JMTM 2015 Jan;4(1):35-38. [doi: 10.7309/jmtm.4.1.8]
4. World Health Organization. Atlas of eHealth country profiles 2015: The use of eHealth in support of universal health coverage. Geneva: World Health Organization; 2015.
5. International Telecommunications Union.: Broadband Commission for Sustainable Development; 2017. Digital Health: A Call for Government Leadership and Cooperation between ICT and Health URL: https://www.broadbandcommission.org/Documents/publications/WorkingGroupHealthReport-2017.pdf [accessed 2018-12-24] [WebCite Cache ID 74tznp9pq]
6. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
7. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
8. Frehywot S, Vovides Y, Talib Z, Mikhail N, Ross H, Wohltjen H, et al. E-learning in medical education in resource constrained low- and middle-income countries. Hum Resour Health 2013 Feb 04;11:4 [FREE Full text] [doi: 10.1186/1478-4491-11-4] [Medline: 23379467]
9. Attwell G. Evaluating e-learning - A guide to the evaluation of e-learning. In: Evaluate Europe Handbook Series. Bremen, Germany: Perspektiven-Offset-Druck; 2006:46.
10. Moodle. URL: https://moodle.org/ [accessed 2019-01-07] [WebCite Cache ID 75FZkkWmG]
11. The Advanced Distributed Learning Initiative. URL: https://www.adlnet.gov/scorm/ [accessed 2019-01-07] [WebCite Cache ID 75FaUnazS]
12. 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]
13. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The Effectiveness of Blended Learning in Health Professions: Systematic Review and Meta-Analysis. J Med Internet Res 2016;18(1):e2 [FREE Full text] [doi: 10.2196/jmir.4807] [Medline: 26729058]

14. Nicoll P, MacRury S, van Woerden HC, Smyth K. Evaluation of Technology-Enhanced Learning Programs for Health Care Professionals: Systematic Review. J Med Internet Res 2018 Apr 11;20(4):e131 [FREE Full text] [doi: 10.2196/jmir.9085] [Medline: 29643049]
15. Bergström S. "Non-physician clinicians" in low income countries. BMJ 2011 May 13;342:d2499. [doi: 10.1136/bmj.d2499] [Medline: 21571913]
16. Ellaway R. Weaving the 'e's together. Med Teach 2006 Nov;28(7):587-590. [doi: 10.1080/01421590600909070] [Medline: 17594547]
17. Cook DA, Ellaway RH. Evaluating technology-enhanced learning: A comprehensive framework. Med Teach 2015;37(10):961-970. [doi: 10.3109/0142159X.2015.1009024] [Medline: 25782599]

Abbreviations

CONSORT: consolidated standards of reporting trials
eHealth: electronic Health
IT: information technology
mHealth: mobile Health
NOS-E: Newcastle-Ottawa Scale for Education
TAM: technology acceptance model

Edited by G Eysenbach; submitted 14.10.18; peer-reviewed by R Vadde, C Jones; comments to author 06.11.18; revised version received 11.11.18; accepted 24.11.18; published 09.01.19.

Please cite as:
Barteit S, Jahn A, Banda SS, Bärnighausen T, Bowa A, Chileshe G, Guzek D, Jorge MM, Lüders S, Malunga G, Neuhann F
E-Learning for Medical Education in Sub-Saharan Africa and Low-Resource Settings: Viewpoint
J Med Internet Res 2019;21(1):e12449
URL: http://www.jmir.org/2019/1/e12449/
doi:10.2196/12449
PMID:30626565

Viewpoint

Use of the Principles of Design Thinking to Address Limitations of Digital Mental Health Interventions for Youth: Viewpoint

Hanneke Scholten[1*], MSc; Isabela Granic[1*], PhD

Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

[*]all authors contributed equally

Corresponding Author:
Hanneke Scholten, MSc
Behavioural Science Institute
Radboud University
PO Box 9104
Nijmegen, 6500 HE
Netherlands
Phone: 31 24 3612527
Email: h.scholten@bsi.ru.nl

Abstract

Numerous reviews and meta-analyses have indicated the enormous potential of technology to improve the appeal, effectiveness, cost, and reach of mental health interventions. However, the promise of digital mental health interventions for youth has not yet been realized. Significant challenges have been repeatedly identified, including engagement, fidelity, and the lack of personalization. We introduce the main tenets of design thinking and explain how they can specifically address these challenges, with an entirely new toolbox of mindsets and practices. In addition, we provide examples of a new wave of digital interventions to demonstrate the applicability of design thinking to a wide range of intervention goals. In the future, it will be critical for scientists and clinicians to implement their scientific standards, methods, and review outlets to evaluate the contribution of design thinking to the next iteration of digital mental health interventions for youth.

(J Med Internet Res 2019;21(1):e11528) doi:10.2196/11528

KEYWORDS

anxiety; depression; design thinking; e-mental health; youth

Background

The prevalence of mental health problems has significantly increased among children (aged 5-10 years) and adolescents (aged 10-24 years [1]) [2-5], and the current prevalence rate of mental disorders is estimated to be 13.4% [2-5]. According to the latest update from the World Health Organization [6], half of all mental health disorders in adulthood start by the age of 14 years and three-quarters, by the mid-20s. As a result of the increasing overall prevalence rates, increases in the rates of mental health concerns for young people specifically [1], and a constant number of available treatments over time, 64%-87% of mental health issues in young people are undetected and untreated [6,7-9].

The rapid growth of technological innovations has been welcomed as an unprecedented opportunity to address the increasing gap between demand and supply of mental health services by many in the mental health research and practice communities [10,11]. Several reviews have indicated the enormous potential of technology to improve the effectiveness, efficiency, cost, reach, personalization, and appeal of mental health interventions [12-16]. Under the rubric of "e-mental health," such advantages are proposed to rely on the ubiquitous role of interactive media in the daily lives of young people [17]. However, the effectiveness of digital technology in reducing the burden of mental health at a population level is progressing slowly at best [18-21]. It remains uncertain whether the hype and promise of e-mental health solutions will actually be realized [15,22].

In the current paper, we summarize and critique the available evidence on the efficacy of digital interventions for young people, with a focus on those targeting anxiety and depression (treatment and indicated prevention). There are many excellent reviews and meta-analyses summarizing the efficacy of digital interventions in the literature; this viewpoint does not attempt to do the same. Instead, we briefly summarize the evidence indicating poor outcomes, especially for youth. We have only included effects based on postintervention measurements,

because most meta-analyses did not have the power to reliably conclude the effects on follow-up measurements. Thereafter, we outline an altogether new approach that has the potential to dramatically improve digital tools for youth mental health by using principles from the discipline of design. Design thinking (DT) is usually considered outside the purview of scientific research; however, we argue that this cross-disciplinary approach may be key to galvanizing progress. Three tenets of DT are introduced here, and preliminary empirical evidence from our own laboratory demonstrates both the opportunities and challenges of this new approach. Finally, as this design framework is new in the mental health arena, we end this paper with recommendations for systematic programs of research that directly test its impact on effectiveness and its implications for implementation.

Outcome Research on Digital Mental Health Interventions

The widespread availability of digital technology has led to a proliferation of digital mental health (DMH) interventions. A substantial part of DMH approaches target depression and anxiety and are based on cognitive behavioral therapy (CBT), originally developed several decades ago for face-to-face treatment and adjusted for self-help books and manuals [12]; we focus our review and recommendations on this class of DMH interventions. There is also a growing body of promising research on virtual reality interventions, especially in the clinical context [23], but it is not directly relevant to our current purposes.

Overall, the efficacy of DMH programs for depressed and anxious adults has been established by over 100 studies [12]. Based on the latest meta-analysis for anxiety disorders among adults [16], guided DMH interventions are more effective than waiting list, attention, information, or online discussion groups. For depression [24], DMH interventions were favored over waiting lists only. There are limited data available to compare DMH interventions and active treatment or placebo control groups; however, the available data suggest that DMH interventions are more effective than other interventions (with smaller effect sizes than those of a waiting list). Other studies also suggest that guided DMH interventions do as well as active treatment control groups (ie, face-to-face CBT) for anxiety [16] and depression [13].

There is still considerable variability in the outcomes, and the inclusion (or exclusion) of human guidance could be one of the key factors that explain this variance. The influence of human guidance on DMH interventions for adult anxiety and depression is still debated [25,26]. Some meta-analyses report equal effects of guided and unguided DMH interventions for anxiety [16] (but with very low-quality evidence) and depression [24], whereas other studies have shown that guided DMH interventions outperform unguided interventions [27-29]. Based on a recent meta-analysis on depression symptoms [30], unguided DMH interventions are more effective than waiting list, attention placebo, no treatment, or treatment as usual; however, its effect is much smaller than that of guided DMH interventions. Importantly, Ebert and Baumeister [25] argue that these meta-analyses are solely based on randomized controlled trials (RCTs), which require a high level of commitment and adherence from patients, unlike the conditions in routine clinical care. It is likely that the reported effect sizes of unguided DMH interventions under laboratory settings are overestimated for their potential in routine clinical care. Thus, despite optimism about the potential of DMH tools as standalone interventions for adults, human involvement in its delivery and monitoring seems to be an important mediator of success for routine clinical care outside the laboratory.

The results are less optimistic for the comparably smaller body of evidence available for children and adolescents. Based on the most recent meta-review by Hollis and colleagues [15], overall effect sizes for youth are moderate to large for anxiety disorders and small to moderate for depression disorders when DMH approaches are compared to a waitlist group. However, analyses comparing DMH interventions for both anxiety and depression with active nontherapeutic controls have generally failed to show superiority of DMH interventions [31,32]. Noninferiority trials comparing DMH interventions to face-to-face CBT in order to determine whether a new intervention is therapeutically similar to an existing effective treatment [33] showed that most DMH approaches were as effective as face-to-face CBT [31,34-36]. In contrast, Pennant and colleagues [32] showed that face-to-face CBT was more effective than DMH interventions. Thus, although the available evidence is not yet conclusive for youth-focused studies, some human guidance seems to be important for the effects on anxiety and depression.

The role of human guidance in youth-focused studies is difficult to ascertain, as the level of human support is poorly specified across trials [15]. However, one of the most recent studies [37] reported that the interventions in their meta-analysis that favored the DMH intervention group included face-to-face guidance, monitoring of engagement, or follow-up telephone calls by teachers and health professionals. Importantly, the quality of the youth-focused studies for DMH programs is generally low to moderate, which is most often a result of methodological flaws such as intervention heterogeneity in terms of content, dose, settings, or quality [15,38]. Furthermore, problems such as insufficient search processes of the literature, small sample sizes, differences at baseline in study samples, and publication bias play important roles in the quality of these studies [32,39]. In summary, there is reasonable support for the role of DMH tools in improving anxiety and depression problems in adults, but there are fewer promising results for children and adolescents. In youth-focused research, poor overall outcomes, heterogeneity of results, and poor quality of many studies prevent definitive conclusions.

Limitations of DMH Interventions for Youth

Despite some promise, there is growing consensus about the limitations of DMH approaches, with almost every meta-analysis and systematic review (both with adult and youth samples) highlighting the same problems. First, high attrition rates and low adherence to protocols are consistently problematic,

especially in unguided interventions [30,37,39,40] as compared to guided DMH interventions [37,41]. Considering only unguided digital interventions, a meta-analytic study by Karyotaki and colleagues [42] showed that almost 70% of participants dropped out before completing 75% of the intervention. Attrition and low adherence are even bigger challenges among children and adolescents; the younger the participant sample, the greater the dropout rate [37,42]. Välimäki and colleagues [37] showed that young people (between the ages of 10 and 24 years) in the digital intervention groups (most often, guided DMH interventions) left the study earlier than the control group participants. Thus, the use of guided digital interventions seems to be the best solution; however, the need for therapists compromises the often espoused advantage of DMH interventions—their scalability (ie, easily deployed across the globe to populations with different economic and ethnic backgrounds [15]).

Most importantly, DMH tools do not remotely approximate the level of attractiveness and interactivity to which young "digital natives" have grown accustomed [17,32,43]. In the first generation of DMH tools, most researchers and intervention scientists seem to have assumed that moving content online and providing youth the agency to navigate this content at their own pace and in their own context makes the content more engaging than that in conventional treatment approaches. However, in a vast majority of cases, the content of DMH interventions is not significantly changed from the manuals from which they were derived. In the understandable and commendable effort to remain "evidence-based," most DMH interventions for youth are a little-more-than modified and uploaded CBT manuals and workbooks (eg, MoodGym, Cool Kids, Camp Cope-a-lot, and BRAVE; for a review discussing the evidence base of these DMH interventions, see [32,44]). It is likely that the digital incarnations of these CBT interventions are rendered even less engaging than their original format because they are less flexible and personalized. No therapist is available to maintain motivation for change, build trust and hope, and sensitively tailor the treatment to personal idiosyncrasies.

Many DMH interventions are based on a one-size-fits-all approach (eg, a linear progression with content released to all participants using time-based rules [44]), which has its advantages because it is systematic, but remains problematic because of its perceived inflexibility [45]. Young people, in particular, value self-reliance and control when accessing digital products [46] or mental health services [47], and current DMH interventions are often perceived as impersonal and unresponsive to their individual needs [15,43]. In addition, DMH interventions are content focused (ie, CBT techniques) and not user focused (ie, they are not designed around how and when young people prefer to engage with digital experiences) [48], resulting in a large disconnect between the world in which youth live and the content and style of DMH interventions. For the current and upcoming generation of youth who play video games and socialize online daily, the norm is digital experiences that are exquisitely designed to adjust to the pace, content preferences, and skill levels of their users [49]. Personalization is consistently mentioned as one of the biggest advantages of digital solutions, but personalization, dynamic adjustment, and tailoring have not been realized with DMH tools thus far [15,43,50].

The cognitive load of DMH programs seems to be an additional limitation especially for young people. Many e-mental health programs are overly pedantic, didactic, and cognitively focused [44], thereby potentially overloading children and youth who find this approach too difficult and inaccessible [51]. Homework assignments pose an additional problem, as they rely on the abilities of the child or adolescent to practice the CBT-based exercises and learn accordingly. Youth very often fail to adequately follow through on these offline homework assignments, because they simply do not understand them well enough to practice or are not motivated to do so [35,39]. The same practice and homework problems can arise in conventional CBT, of course, but in face-to-face treatments, therapists are present to motivate, encourage, answer questions, and keep clients accountable [12].

At this point, an important caveat is in order: We are by no means advocating exclusion of therapists, coaches, and teachers altogether, especially in serious, chronic mental health cases among youth. Our best outcomes for serious clinical youth cases may come from combining face-to-face interactions with digital intervention "homework," in which young people practice the lessons they have learned in the comfort of their own home or on mobile devices embedded in their everyday lives (eg, "blended" approaches [52]). However, this digital homework still requires attention to be paid to the factors that motivate and engage users. For less severe mental health cases, DMH programs may serve as stand-alone interventions preventing at-risk youth from symptom aggravation.

In summary, a convincing set of reviews and meta-analyses suggest that the promises of digital solutions, especially those targeted at youth, have not yet been realized [11,12,37]. Specifically, the benefits of DMH interventions, including increased engagement and motivation, fidelity to intervention protocols, and opportunities for personalization [12,48,50,53], remain largely unrealized. Perhaps unsurprisingly, all the reviews we have summarized end with general recommendations for reflection and reform, urging future efforts to take engagement, retention, and fidelity more seriously. However, these critical reflections consistently stop at that point, providing no concrete, actionable solutions to address the limitations they revealed [54-57]. In the rest of this viewpoint, we elaborate on a set of guiding principles and concrete strategies to potentially address this impasse.

Design Thinking: Novel Recommendations and Proposed Solutions

In the following section, we outline a design framework that has helped us reimagine the development of DMH solutions for children and adolescents. Our approach started by identifying the limitations of past DMH interventions for youth and attempting to directly address each of them. A major step toward such solutions derives from our work with applied video games for mental health. We previously provided a detailed empirical

review [49] that supports the rationale for using digital games as intervention tools for young people. In short, well-designed applied games are intrinsically motivating, offer a strong sense of agency, and are simply fun. They also provide a compelling virtual playground to not only gain knowledge, but also *practice* skills. Finally, applied games can overcome the stigma associated with traditional and self-help interventions.

We are not the first to suggest that applied games are useful intervention approaches. A zeitgeist has emerged in the medical and educational fields for applied or "serious" games as tools for enhancing medical care [58-61]. Although much less work has been done with serious games for mental health as compared to other conditions, several game-based interventions have been developed [62-65]. A large part of our message is that not all digital interventions, including games, are designed equally, and most serious games have the same general limitations that we outlined for DMH interventions. Our solution has been to adopt a DT framework, which provides a general cohesive set of principles and recommendations for DMH delivery.

Defining Design Thinking

Before defining DT, it is important to understand that this approach is not simply about making products or services more attractive, pretty, or graphically sophisticated. It often involves some degree of esthetic improvement, but fundamentally, DT is both a mindset and a set of practices that are solution based. The business community as well as the healthcare, transportation, and creative industries have benefited enormously from the adoption of DT [56,66-70]. Compared to scientific practices in which data are "objective," observable facts that are tested against *a priori* hypotheses, DT is a fundamentally subjective practice that focuses on discovering the emotional needs of users, their idiosyncratic contexts, their motivational concerns, and other related entities. DT aims to build a practical product or service that serves a very specific need.

We do not suggest that DT is enough as a stand-alone practice to address the concerns we have listed about DMH interventions for youth. However, combined with rigorous scientific standards and methodologies, this cross-disciplinary approach holds a great deal of promise. There are three core tenets of DT [67,68]: Empathy, which is a human-centered approach that keeps the emotional, motivational, and functional needs of users at the center of the development process; Multidisciplinary Ideation, which involves solutions generated by cross-disciplinary teamwork and collaboration; and Experimentation, which is the practice of rapid prototyping and iteratively testing products or services with target users during, rather than after, the development phase. These terms have varied meanings in psychology, psychiatry, and clinical practice, but have very specific meanings in the discipline of design, as elaborated in the next section.

Empathy

At its core, empathy-based design is a human-centered approach that answers the question "who is it for?" rather than "what does the product look like and contain?" Empathy seems like a fuzzy, unscientific lens through which evidence-based practice is considered, but it is the most crucial and, perhaps, least understood or integrated practice in the development of digital interventions. Empathy-driven design seeks to optimize user engagement, immersion, and motivation and as such, it has the potential to address key limitations of conventional DMH approaches (ie, high attrition, low adherence, and cognitive load). Instead of starting with the common premise, "we're going to design an app that does X," empathic design begins with "we're going to solve X for a specific population" and thus helps developers expand beyond the exclusive content focus of DMH programs (eg, CBT techniques) toward user concerns (eg, a young person's preferences and digital habits).

Beyond understanding the demographics, personalities, and preferences of individual users, empathic design keeps the whole end-to-end user *experience* in mind. Applied to youth mental health, user experience can be conceptualized according to these questions: (1) How are young people going to find an intervention, game, or service? (2) When they find it, does that digital ecosystem motivate them to keep discovering more, or does it shut them down? (c) Are there positive expectations for change embedded in a growth mindset? [71,72] (3) How long after they purchase or freely download the product, service, or game will it provide feedback about progress, and how will that make users feel? (4) Can they share it with like-minded peers and concerned adults? (5) Will it be updated with new content to keep them interested over longer periods? (6) When the experience ends, is there a feeling of mastery?

An empathy-driven approach also includes *participatory design*: We not only design *for* young people but *with* them as well and do so from the start of the design process. As digital natives [73], young people are using interactive media and technology almost from birth [17] and on a daily basis. By the time they engage with any particular DMH product, they have grown accustomed to interacting with highly engaging, sophisticated, and immersive contexts. If digital interventions are to stand a chance of improving the mental health of youth in the coming decades, they will need to be designed to stimulate and retain users' attention. The first step towards ensuring that this will happen is to invite these users to codevelop products aimed at their cohort. Several other researchers have suggested the importance of recruiting young people in the development process. This practice has been referred to as participatory design, participatory research, codesign, and user-centered design [43,62,63,74,75]. However, in the mental health context, this process often amounts to professionals asking youth about the products they have already designed, with little time or money allocated to the suggested changes that emerge through the process [74,76,77].

We argue that, in the mental health fields, the greatest barrier to adopting a participatory approach, is the implicit paternalistic mindset that may have become ingrained in many academics and practitioners. Mental health researchers and clinicians often assume that young people, especially those who are emotionally vulnerable, do not know when they are suffering and are incapable of asking for the kind of help they need [7,78,79]. However, most youth with anxious and depressive symptoms are well aware of their vulnerabilities and struggles [80]. The key barrier to improving outcomes for these youth is not their own ignorance of whether they need help or even the kind of

help they need, but their ability to find the resources and services that will support and train them in a way that speaks to their preferences and modes of learning [78,80,81]. By recruiting youth with mental health challenges from the outset of the design process in order to teach us how they interact and seek information online, we have a better chance of designing interventions that they will find initially engaging, will retain their attention, and will ultimately be viewed as relevant to their needs [46].

In our work, we have taken on this empathy-driven, participatory approach to fundamentally change our starting point in applied game design. Participatory design starts very early, even before any programming of intervention games has begun, by using paper prototyping methods, interviews, and focus groups. Traditionally, this user-research phase is rushed through to reach the "real science." However, we argue that the scientific outcomes we seek to enhance will not be realized unless empathic, participatory methods are placed at the forefront of our process.

A specific example from our laboratory may be useful to clarify the advantages of using an empathy-driven approach for digital intervention design. In this project, we aimed to design and test a game to help young people quit smoking. Although this game did not directly target mental health, the example is illustrative of the principles and practices that are entirely relevant to mental health applications and useful for clarifying the practices we previously described in the Empathy section. This example is especially relevant to mental health, as the design of this applied game capitalizes on the social peer structure of youth. We tried to increase motivation, commitment, and engagement through game-based experiences that were fundamentally interactive and brought them together with like-minded peers. These social processes are equally important in the design of interventions for anxious and depressed youth, as social ties play a beneficial role in maintaining psychological well-being and mental health [82,83].

Before designing this applied game, we invited young people who smoke to talk specifically about their smoking experiences and how they feel about quitting. Past research on smoking cessation claims that because young people have only just started smoking, they are not motivated to quit [84]. Thus, psychoeducational programs that outline the negative consequences of smoking are the most common intervention approach; skill training based on CBT techniques and motivational interviewing are traditionally employed as well [85-87]. However, none of these interventions have been successful [85,87].

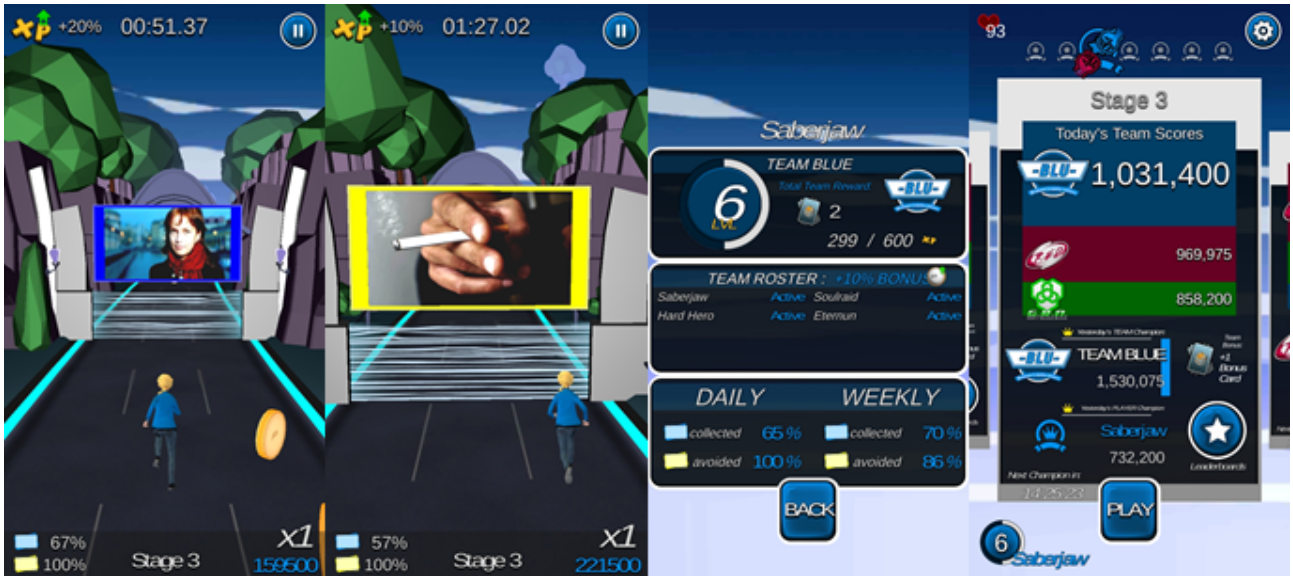
We used a different approach by using structured tools from DT with young people who smoke (eg, card-sort tasks, screen-shot photos of youths' own phones, and interview protocols; d-school resources [88]). The insights we reviewed were all gained from qualitative interviews with young smokers that were part of unpublished user research with early versions of the game. Our aim was to discover previously misunderstood or overlooked factors that could contribute to designing an end-to-end intervention experience that would effectively help youth quit smoking. We learned that contrary to common assumptions, youth are well aware of the negative consequences of smoking and are often motivated to quit [89-91]. Despite their motivation to quit, they did not know where to look for help; there are no evidence-based interventions available thus far for young people attempting to quit [85,87]. They explained their feelings of inferiority and anxiety that accompany failed attempts to quit. Considering the stigma associated with smoking, they resist asking for help with their addictive vulnerabilities, at least from adults. They are aware that they are struggling and some seek help online anonymously. However, the advice they receive online is perceived as didactic, outdated, and boring.

Instead of focusing on the unhealthy and harmful outcomes of smoking, an empathy-driven lens led us to delve deeper into the emotional and social contexts from which smoking behaviors emerge. We attempted to understand what smoking meant to these young people, how it served important needs, and where they felt that smoking blocked their goals. We discovered that there was a great deal of variability in terms of where and when young people chose to smoke, suggesting the importance of tailoring a DMH intervention to these individual preferences. We also learned that smoking served several functions: to cope with stress; to overcome boredom during the day (eg, waiting for the bus); and crucially, to socialize with friends during breaks. These functional and motivational accounts of young smokers served as the essential scaffold on which we based other evidence-based practices, such as inhibition training [92].

From these empathy-focused conversations, we designed an intervention to serve as a functional replacement for the smoking habit. We developed the game as a "casual runner," a genre that lends itself to short bursts of intensely engaging gameplay (ie, 3-5 minutes per session, which is the approximate time taken to overcome a craving moment or smoke a cigarette). To address the problems with the one-size-fits-all approach, we ensured that the game could be played during individualized moments of high craving or boredom. We designed tailored prompts that reminded users to play at instances when they reported experiencing high levels of craving. To enhance relevance in youths' everyday lives, the game is played on mobile devices, so that young people had access to it whenever they might want to smoke. Figure 1 presents screenshots and a leaderboard example.

We learned how important it was to bring their peer network into the intervention context. We brought them together with like-minded peers who smoked but were motivated to quit through cooperative (and competitive) team-based gameplay that mimicked other online social games with which they were already familiar. Through the cooperative team-based design, youth could learn that there were many like-minded peers that experience the same problems they do, and they could playfully apply "friendly" peer pressure to encourage each other to play the game, which implicitly indicated that they were all quitting together. The competitive elements helped them stay motivated and focused on quitting without resorting to didactic or stigmatizing scare tactics.

Figure 1. Screenshots of HitnRun showing the runner game and leaderboards.



RCTs to test the efficacy of this new approach and whether the specific design elements mediate efficacy are underway. We do not have these data yet. However, our main aim of elaborating this example was to provide concrete instantiations of design decisions that would not have otherwise emerged without an empathy-based DT approach.

Multidisciplinary Ideation

DT places immense value on cross- and interdisciplinary collaborations with the conviction that true innovation can only arise through a multiplicity of perspectives. A crucial part of DT practices is the generation of a large set of ideas without evaluating the veracity of those ideas in the initial phases and simply collecting the broadest range possible. This approach is in stark contrast to the approach most scientists take, starting from a place of established principles and evidence-based techniques. Although we strongly believe that scientific principles and practices should form the basis of DMH interventions, the potential for new opportunities to engage and retain young people's attention and time may stem from allowing teams to creatively explore options outside of these empirically established methods. Such exploration is much more likely to yield genuinely novel design possibilities when diverse perspectives are encouraged and then culled via scientific constraints.

We suggest a wide multidisciplinary approach to use DT practices for the development of immersive digital products for young people's mental health. For example, in our work, we cultivate collaborations among developmental psychologists, neuroscientists, veteran game developers who have extensive experience in the commercial game industry, programmers, and artists, all of whom need to learn each other's domain-specific language and codevelop a set of shared terms and goals. For the DT methods to work, it is important to invite stakeholders such as teachers, clinicians, physicians, parents, and children themselves to be a part of the codesign process. Through this approach, we can integrate empirically validated principles of clinical change with evocative art and design to render user experiences that are enriching, engaging, and "sticky" enough to bring young people back for more.

The application of the DT framework to DMH interventions requires ideation from more than designers, artists, and mental health professionals. Programmers and formally trained engineers are also crucial partners. Most often, technology and technical requirements are ignored by social scientists. However, early and frequent collaborations with engineers during the early design and evaluation stages are critical, because this is when the back-end, data-acquisition system can be seamlessly integrated with the front-end user interface. This back-end architecture can prove incredibly useful for researchers and clinicians alike. For example, strong, effectively designed data-acquisition systems can be designed to automatically calculate and quantify real-time in-game (or in-app) play or usage behavior. Information about what parts users interacted with, how long they engaged, when they returned, in what areas they lingered longest, how quickly they acquired skills, and other such parameters can serve as powerful analytic tools for the researcher and clinician. Thus, engineers who can build analytic, noninvasive systems can address some of the most pernicious limitations of conventional DMH interventions: participant and client accountability, fidelity, and tracking. In addition, technical experts need to be involved beyond the development and efficacy testing in order to update software continuously (to keep it current and more engaging) and ensure compatibility with changing technology ecosystems (eg, new operating systems and various platforms such as phones, watches, and tablets).

Experimentation

"Design thinking is a misnomer; it is more about doing than thinking." [88]. It seems peculiar to explain experimentation to researchers, but in the context of DT, the meaning of experimentation is different from applying a scientific method in a controlled environment in which one, or very few, factors are manipulated to test a hypothesis. Experimentation in DT refers to a set of processes and practices built around prototyping. A *prototype* is a simplified version of a product,

or part of a product, that is created in minimal time and at minimal cost. It is used to test the validity of ideas or design assumptions as rapidly and cheaply as possible. Designers often emphasize the massive advantages of "just doing" (ie, acting out ideas to test their utility before a great amount has been invested in a product or service). In the case of DMH interventions, this prototyping phase is often skipped or applied at such a late stage that only little adjustment is feasible.

Prototyping takes various forms (eg, paper-and-pencil games, whiteboards with sticky notes that depict the flow of a digital experience, storyboards that illustrate the "beats" of a user's end-to-end experience, and presentation mock ups to click through to get the feel of a tool). All these forms are concrete, tangible artifacts that allow hands-on experience and evaluation before any programming starts and are usually applied iteratively with a small number of target users. Importantly, prototyping is not meant to replace scientifically rigorous experiments or clinical trials; rather, it addresses specific design questions. The results of prototyping iteratively and rapidly are action-based insights about the feel and usability of a product. Often, what emerge are "creative serendipity" and unanticipated insights.

One of the most important lessons we have learned is that throughout the experimentation process (prototyping and later phases), it is crucial to separate and synchronize the goals related to the digital tool versus the intervention. We have studied game design, in particular, and we will focus on that domain for articulating our points, but the same principles apply to any interactive app, dynamic website, or other digital media form. The timelines for game development and intervention development run in parallel (Figure 2). Importantly, these streams iteratively influence one another over time. Each domain has its own set of testing principles and practices that are to be applied differentially at each phase.

For example, early in the prototyping phase, two sets of goals are evaluated in parallel (Figure 3). On the game-development level, we evaluate whether the game's mechanics (ie, "verbs" of the game) actually work as they were designed (what players actually do to move through the game towards specified goals; this could also be navigation procedures for a website or app). At this early stage, we test whether players proceed through the intended pathways. Do they know what to do to solve a puzzle? Do the controls feel natural? Concomitantly, on the intervention-development level, this phase is often referred to as piloting and can include tests of whether the game elicits the emotional responses intended. Is the cognitive load overwhelming (a barrier with conventional digital interventions for youth)? Do players respond with reactance (ie, backlash or negative affect experienced in response to unsolicited advice)? Do they experience the game as didactic or pedantic and quickly turn it off? Figure 3 also shows the relation between the scope of data collection (eg, sample size), the timing of evaluations, and the different foci and products over the course of the development process. All the prototyping and testing discussed so far fall under Box A in Figure 3. An example from our laboratory with an applied game that has undergone most of the phases in Figure 3 is presented next.

Example: MindLight, an Anxiety-Prevention Game for Children

During the development of MindLight, a game designed to decrease anxiety symptoms in children, a great deal of prototyping was performed to address the two streams of design goals. For example, the game relied largely on exposure techniques to train anxious children to practice facing fears while using relaxation methods. The artists on the project drew several versions of the monsters in the game (Figure 4), given the importance of these figures for triggering fear, and tested whether children would approach them after a certain period of hesitation (game-development goal; Figure 3). The psychologists on the project tested children's fear responses and appraisals of control to overcome their fear of each of these creatures (intervention-development goal). Contrary to the expectations, most children found the one-eyed monsters humorous and "cute." Thus, we chose to use a two-eyed creatures instead, to ensure we triggered the fearful responses essential for exposure techniques to work at the intervention level.

Another example of a game mechanic that needed repeated prototyping was neurofeedback. We designed the game so that the calmer children felt while using relaxation techniques during exposure to fearful events (measured by a one-channel electroencephalography system [93,94]), the brighter the light in the game would shine; the more anxious the child felt, the more the light dimmed. A sensitively tuned threshold for when the light would turn on had to be established: players needed to feel motivated when it was dark to practice relaxation skills and maintain motivation to regain their calm, but they could not be so afraid or frustrated that they quit early. Pilot studies helped us identify this threshold as well as a reasonable pace of increasing the threshold over the course of the game while maintaining challenge and engagement.

After several iterations to tweak the dynamic adjustment and reward system, a redesigned beta version (full game coded with 8 hours of gameplay) was used in a series of RCTs. During this phase (D in Figure 3), the main intervention goals were to use rigorous experimental designs to test the game's impact on children's anxiety symptoms. Results from four RCTs were reassuring: The data consistently showed significant decreases in children's anxiety symptoms, with two of the studies showing similar improvements as active control [94] and treatment -as-usual [95] studies and two studies showing improvements equivalent to cognitive-behavioral interventions [93,96], even after long-term follow-up [93].

At the same time, we tested critical elements at the game-development level, including replay ability, engagement, and the likeliness that children would recommend the game to others. Data showed that children were equally likely to recommend MindLight to a friend as one of the most popular commercial games for this age group [94], and they consistently rated the game as fun and engaging [93], suggesting that our prototyping phase was successful. Mediation studies were also performed to examine whether the training mechanics that were designed based on evidence-based techniques (eg, exposure and light-based neurofeedback) were the action mechanisms that explained outcomes and determine if the results confirmed our

hypotheses [97]. As expected, children reported feeling fearful of the monsters in the game (ie, exposure worked). More importantly, *increases* in children's capacity to shine their "mindlight" (the light indicating relaxation measured with neurofeedback) across game sessions predicted reductions in anxiety symptoms 3 months later.

Figure 2. Separate but interactive development timelines for game and intervention goals, with more frequent testing and iterative prototyping at the start of the process than at the end.

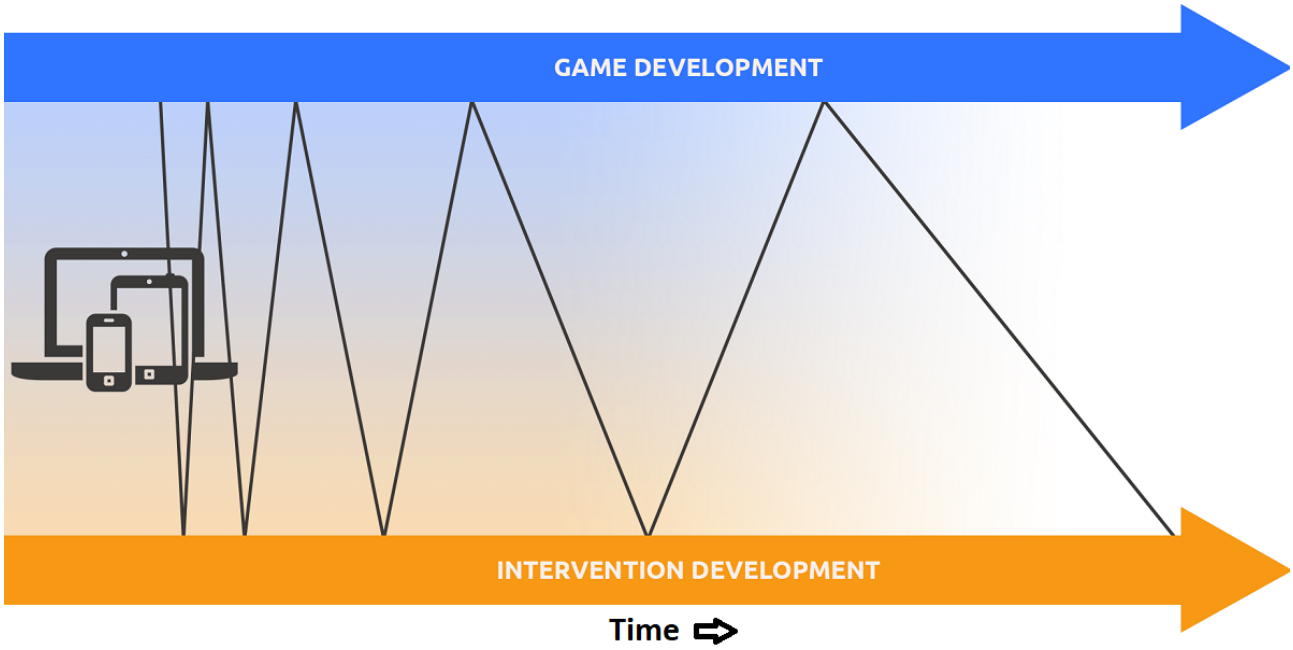


Figure 3. Interaction of the timeline and scope of game development with intervention development. RCT: randomized controlled trial.

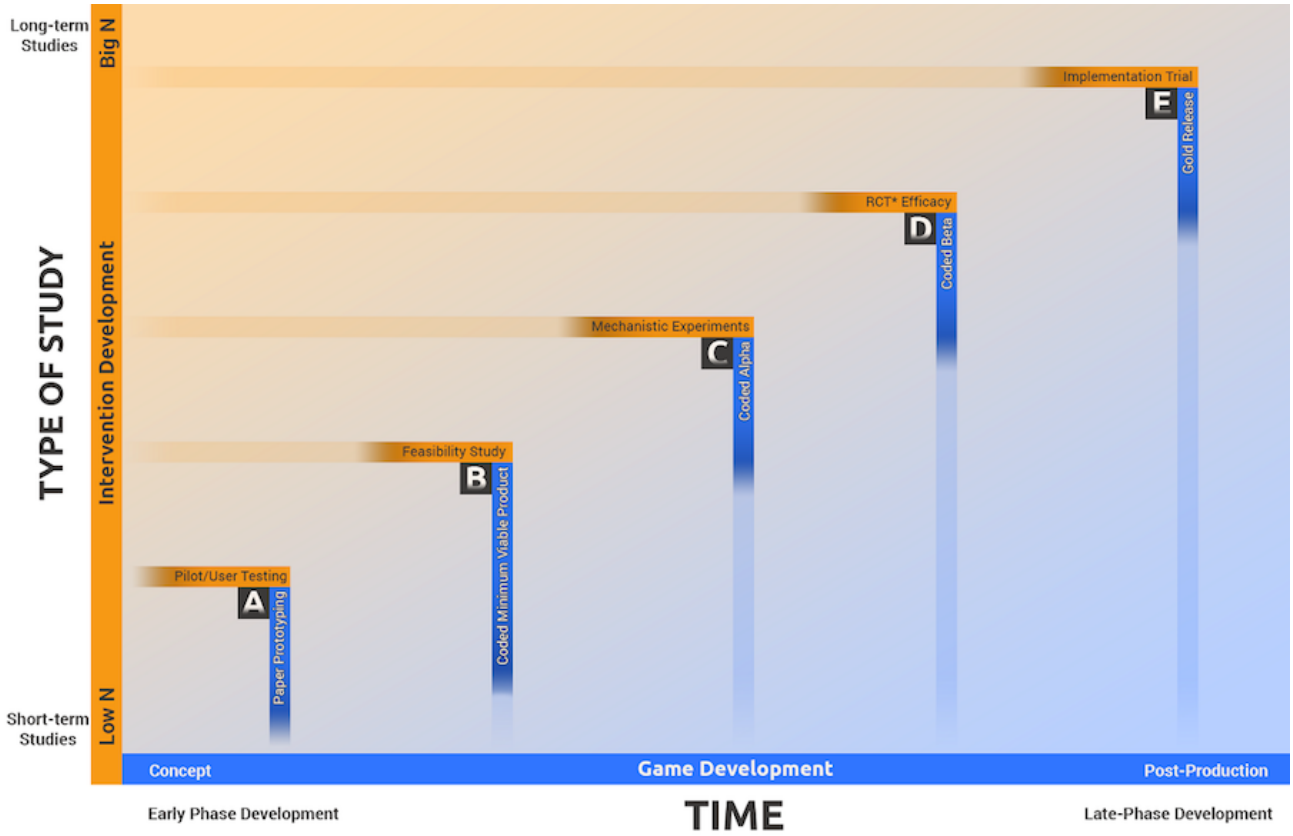


Figure 4. Cat concept design for MindLight.

We presented MindLight in this paper as an example to illustrate how the framework in Figure 3 can be concretely applied to the development and research of a digital intervention tool for mental health. We also attempted to highlight how the DT framework was integral to developing an effective digital anxiety intervention tool that was eagerly played by children repeatedly. Importantly, this framework should be applicable to a wide range of digital interventions and is certainly not restricted to applied game development.

Implementation Considerations

The last stage of Figure 3 is the implementation phase with the "gold release" version of the DMH product, adjusted with insights from the RCTs and mediation studies and polished for distribution purposes. Related to the second DT tenet, rolling out DMH interventions requires a multidisciplinary effort [98]. For digital tools, in particular, we may need to engage more people than stakeholders and policy makers and consider the unique expertise of marketing experts, business leaders, and technical support teams. There are crucial issues to be considered with commercially oriented partners (eg, scientific integrity and conflicts of interest), but if scalability and broad impact are the aims, marketing and business experts may be key to developing optimal models of service delivery. In this final stage, at the game-development level, it is important to consider whether young people discover the games we develop on their own through their own online search initiatives; whether they are interested in engaging with our content; whether we can retain that attention and motivate them to practice skills; and the extent to which they share these DMH programs with peers and family that might benefit similarly from them. On the intervention- development level, implementation tests may need to go beyond RCTs. Current technologies rapidly change in a few years, and there is no reason to believe this rate of change will slow down. In the midst of this rapidly shifting technological landscape, the traditional research designs that require interventions to remain stable across many years may be less practical, useful, and feasible [48,99,100]. Researchers are reconceptualizing the scientific framework, methodology, and implementation strategies that might better suit implementation and outcome studies in the DMH context [48,99,100]. DT and its evaluation practices, with their focus on qualitative and participatory studies, seem to have some useful recommendations in this regard.

Conclusions

Several reviews have indicated the enormous potential of technology to improve effectiveness, efficiency, cost, reach, personalization, and appeal of mental health interventions for young people. However, significant challenges including engagement, retention, fidelity, lack of personalization, and cognitive load continue to hinder progress in this field. Thus far, all meta-analyses and reviews have highlighted these barriers but have not offered any avenues for actionable solutions [54-56]. We introduced three tenets of DT—empathy, multidisciplinary ideation, and experimentation—and showed how these mindsets and practices can inform the development of future digital interventions. We also provided concrete examples from our work to demonstrate how this new approach can be implemented for young people and provided some preliminary evidence that it can improve outcomes and have an impact on engagement. Ultimately, we argued that integrating DT mindsets and practices with conventional scientific approaches is a promising avenue through which digital tools can address youth mental health. However, we are only at the beginning of merging design and science in the mental health arena. As a discipline, design has been criticized for its lack of quality control, the absence of systems to evaluate the quality, and standardization or documentation of the various DT methods [56,101]. In the future, it will be critical for social scientists and clinical researchers who are interested in appropriating DT to use their scientific standards, methods, and review outlets to evaluate the contribution of DT.

Acknowledgments

This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant agreement no. 683262). We would like to thank Evan Hirsh and Ken Koontz for their valuable contribution to the manuscript.

Conflicts of Interest

None declared.

References

1. Collishaw S. Annual research review: Secular trends in child and adolescent mental health. J Child Psychol Psychiatry 2015 Mar;56(3):370-393. [doi: 10.1111/jcpp.12372] [Medline: 25496340]
2. Mojtabai R, Olfson M, Han B. National Trends in the Prevalence and Treatment of Depression in Adolescents and Young Adults. Pediatrics 2016 Dec;138(6):e20161878 [FREE Full text] [doi: 10.1542/peds.2016-1878] [Medline: 27940701]
3. Olfson M, Druss BG, Marcus SC. Trends in mental health care among children and adolescents. N Engl J Med 2015 May 21;372(21):2029-2038. [doi: 10.1056/NEJMsa1413512] [Medline: 25992747]
4. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry 2015 Mar;56(3):345-365. [doi: 10.1111/jcpp.12381] [Medline: 25649325]
5. Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC. The age of adolescence. Lancet Child Adolesc Health 2018 Mar;2(3):223-228. [doi: 10.1016/S2352-4642(18)30022-1] [Medline: 30169257]
6. World Health Organization. Adolescents: health risks and solutions URL: http://www.who.int/en/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions [accessed 2018-07-09] [WebCite Cache ID 70mKtGujT]
7. Rickwood DJ, Deane FP, Wilson CJ. When and how do young people seek professional help for mental health problems? Med J Aust 2007 Oct 1;187(7 Suppl):S35-S39. [Medline: 17908023]
8. Zachrisson HD, Rödje K, Mykletun A. Utilization of health services in relation to mental health problems in adolescents: a population based survey. BMC Public Health 2006;6:34 [FREE Full text] [doi: 10.1186/1471-2458-6-34] [Medline: 16480522]
9. Zwaanswijk M, Verhaak PFM, Bensing JM, van der Ende J, Verhulst FC. Help seeking for emotional and behavioural problems in children and adolescents: a review of recent literature. Eur Child Adolesc Psychiatry 2003 Aug;12(4):153-161. [doi: 10.1007/s00787-003-0322-6] [Medline: 14505065]
10. Bhugra D, Tasman A, Pathare S, Priebe S, Smith S, Torous J, et al. The WPA-Lancet Psychiatry Commission on the Future of Psychiatry. Lancet Psychiatry 2017 Oct;4(10):775-818. [doi: 10.1016/S2215-0366(17)30333-4] [Medline: 28946952]
11. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Dec;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
12. Andersson G. Internet-Delivered Psychological Treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
13. Andersson G, Topooco N, Havik O, Nordgreen T. Internet-supported versus face-to-face cognitive behavior therapy for depression. Expert Rev Neurother 2016;16(1):55-60. [doi: 10.1586/14737175.2015.1125783] [Medline: 26610160]
14. 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 Dec 08;18(9):e236 [FREE Full text] [doi: 10.2196/jmir.5670] [Medline: 27608548]
15. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
16. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2016 Mar 12;3:CD011565. [doi: 10.1002/14651858.CD011565.pub2] [Medline: 26968204]
17. The Lancet Child Adolescent Health. Growing up in a digital world: benefits and risks. Lancet Child Adolesc Health 2018 Feb;2(2):79. [doi: 10.1016/S2352-4642(18)30002-6] [Medline: 30169237]
18. Apolinário-Hagen J, Kemper J, Stürmer C. Public Acceptability of E-Mental Health Treatment Services for Psychological Problems: A Scoping Review. JMIR Ment Health 2017 Apr 03;4(2):e10 [FREE Full text] [doi: 10.2196/mental.6186] [Medline: 28373153]
19. Batterham PJ, Sunderland M, Calear AL, Davey CG, Christensen H, Teesson M, et al. Developing a roadmap for the translation of e-mental health services for depression. Aust N Z J Psychiatry 2015 Sep;49(9):776-784. [doi: 10.1177/0004867415582054] [Medline: 25907269]
20. Meurk C, Leung J, Hall W, Head BW, Whiteford H. Establishing and Governing e-Mental Health Care in Australia: A Systematic Review of Challenges and A Call For Policy-Focussed Research. J Med Internet Res 2016 Jan 13;18(1):e10 [FREE Full text] [doi: 10.2196/jmir.4827] [Medline: 26764181]

21. Wallin EEK, Mattsson S, Olsson EMG. The Preference for Internet-Based Psychological Interventions by Individuals Without Past or Current Use of Mental Health Treatment Delivered Online: A Survey Study With Mixed-Methods Analysis. JMIR Ment Health 2016;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
22. Wykes T, Brown M. Over promised, over-sold and underperforming? - e-health in mental health. J Ment Health 2016;25(1):1-4 [FREE Full text] [doi: 10.3109/09638237.2015.1124406] [Medline: 26732733]
23. Rizzo AS, Koenig ST. Is clinical virtual reality ready for primetime? Neuropsychology 2017 Nov;31(8):877-899. [doi: 10.1037/neu0000405] [Medline: 29376669]
24. 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 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
25. Ebert DD, Baumeister H. Internet-Based Self-help Interventions for Depression in Routine Care. JAMA Psychiatry 2017 Aug 01;74(8):852-853. [doi: 10.1001/jamapsychiatry.2017.1394] [Medline: 28636717]
26. Karyotaki E, Riper H, Cuijpers P. Internet-Based Self-Help Interventions for Depression in Routine Care-Reply. JAMA Psychiatry 2017 Aug 01;74(8):853. [doi: 10.1001/jamapsychiatry.2017.1406] [Medline: 28636715]
27. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
28. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
29. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
30. 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]
31. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
32. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, et al. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
33. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA 2012 Dec 26;308(24):2594-2604. [doi: 10.1001/jama.2012.87802] [Medline: 23268518]
34. Podina I, Mogoase C, David D, Szentagotai A, Dobrean A. A Meta-Analysis on the Efficacy of Technology Mediated CBT for Anxious Children and Adolescents. J Rat-Emo Cognitive-Behav Ther 2015 Nov 17;34(1):31-50. [doi: 10.1007/s10942-015-0228-5]
35. Rooksby M, Elouafkaoui P, Humphris G, Clarkson J, Freeman R. Internet-assisted delivery of cognitive behavioural therapy (CBT) for childhood anxiety: systematic review and meta-analysis. J Anxiety Disord 2015 Jan;29:83-92. [doi: 10.1016/j.janxdis.2014.11.006] [Medline: 25527900]
36. Ye X, Bapuji SB, Winters SE, Struthers A, Raynard M, Metge C, et al. Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: a systematic review and meta-analysis. BMC Health Serv Res 2014 Jul 18;14:313 [FREE Full text] [doi: 10.1186/1472-6963-14-313] [Medline: 25037951]
37. Välimäki M, Anttila K, Anttila M, Lahti M. Web-Based Interventions Supporting Adolescents and Young People With Depressive Symptoms: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2017 Dec 08;5(12):e180 [FREE Full text] [doi: 10.2196/mhealth.8624] [Medline: 29222079]
38. Das JK, Salam RA, Lassi ZS, Khan MN, Mahmood W, Patel V, et al. Interventions for Adolescent Mental Health: An Overview of Systematic Reviews. J Adolesc Health 2016 Oct;59(4S):S49-S60 [FREE Full text] [doi: 10.1016/j.jadohealth.2016.06.020] [Medline: 27664596]
39. Clarke AM, Kuosmanen T, Barry MM. A systematic review of online youth mental health promotion and prevention interventions. J Youth Adolesc 2015 Jan;44(1):90-113. [doi: 10.1007/s10964-014-0165-0] [Medline: 25115460]
40. 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 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
41. van Ballegooijen W, Cuijpers P, van Straten SA, Karyotaki E, Andersson G, Smit JH, et al. Adherence to Internet-based and face-to-face cognitive behavioural therapy for depression: a meta-analysis. PLoS One 2014;9(7):e100674 [FREE Full text] [doi: 10.1371/journal.pone.0100674] [Medline: 25029507]
42. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]

43. Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014;9(1):e84323 [FREE Full text] [doi: 10.1371/journal.pone.0084323] [Medline: 24465404]
44. Stiles-Shields C, Ho J, Mohr DC. A review of design characteristics of cognitive behavioral therapy-informed behavioral intervention technologies for youth with depression and anxiety. Digit Health 2016;2:2055207616675706 [FREE Full text] [doi: 10.1177/2055207616675706] [Medline: 29942571]
45. Johansson O, Michel T, Andersson G, Paxling B. Experiences of non-adherence to Internet-delivered cognitive behavior therapy: A qualitative study. Internet Interventions 2015 May;2(2):137-142. [doi: 10.1016/j.invent.2015.02.006]
46. Boyd D. It's Complicated : The Social Lives of Networked Teens. New Haven, Connecticut: Yale University Press; 2014.
47. Plaistow J, Masson K, Koch D, Wilson J, Stark RM, Jones PB, et al. Young people's views of UK mental health services. Early Interv Psychiatry 2014 Feb;8(1):12-23. [doi: 10.1111/eip.12060] [Medline: 23773401]
48. 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]
49. Granic I, Lobel A, Engels RCME. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]
50. Shafran R, Bennett SD. Commentary: All that glisters is not gold - reflections on Hollis et al. 2017. J Child Psychol Psychiatry 2017 Apr;58(4):504-506. [doi: 10.1111/jcpp.12715] [Medline: 28346756]
51. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010 Sep;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]
52. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing Online and Face-to-Face Therapy: How to Benefit From Blended Care in Mental Health Care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
53. Musiat P, Tarrier N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. Psychol Med 2014 Nov;44(15):3137-3150. [doi: 10.1017/S0033291714000245] [Medline: 25065947]
54. Hilgart MM, Ritterband LM, Thorndike FP, Kinzie MB. Using instructional design process to improve design and development of Internet interventions. J Med Internet Res 2012;14(3):e89 [FREE Full text] [doi: 10.2196/jmir.1890] [Medline: 22743534]
55. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
56. Orlowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory Research as One Piece of the Puzzle: A Systematic Review of Consumer Involvement in Design of Technology-Based Youth Mental Health and Well-Being Interventions. JMIR Hum Factors 2015;2(2):e12 [FREE Full text] [doi: 10.2196/humanfactors.4361] [Medline: 27025279]
57. Zanaboni P, Ngangue P, Mbemba GIC, Schopf TR, Bergmo TS, Gagnon M. Methods to Evaluate the Effects of Internet-Based Digital Health Interventions for Citizens: Systematic Review of Reviews. J Med Internet Res 2018 Jun 07;20(6):e10202 [FREE Full text] [doi: 10.2196/10202] [Medline: 29880470]
58. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Computers & Education 2012 Sep;59(2):661-686. [doi: 10.1016/j.compedu.2012.03.004]
59. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107. [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
60. Girard C, Ecalle J, Magnan A. Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. Journal of Computer Assisted Learning 2012 Jun 13;29(3):207-219. [doi: 10.1111/j.1365-2729.2012.00489.x]
61. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
62. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious Games and Gamification for Mental Health: Current Status and Promising Directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
63. 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]
64. Lau HM, Smit JH, Fleming TM, Riper H. Serious Games for Mental Health: Are They Accessible, Feasible, and Effective? A Systematic Review and Meta-analysis. Front Psychiatry 2016;7:209 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]
65. Li J, Theng Y, Foo S. Game-based digital interventions for depression therapy: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2014 Aug;17(8):519-527. [doi: 10.1089/cyber.2013.0481] [Medline: 24810933]

66. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. Qual Saf Health Care 2006 Oct;15(5):307-310 [FREE Full text] [doi: 10.1136/qshc.2005.016527] [Medline: 17074863]
67. Brown T. Design thinking. Harv Bus Rev 2008 Jun;86(6):84-92, 141. [Medline: 18605031]
68. Brown T, Wyatt J. Design Thinking for Social Innovation. Development Outreach 2010 Jul;12(1):29-43. [doi: 10.1596/1020-797x_12_1_29]
69. Duncan AK, Breslin MA. Innovating health care delivery: the design of health services. Journal of Business Strategy 2009 Feb 27;30(2/3):13-20. [doi: 10.1108/02756660910942427]
70. Johnson AE, Winner L, Simmons T, Eid SM, Hody R, Sampedro A, et al. Using Innovative Methodologies From Technology and Manufacturing Companies to Reduce Heart Failure Readmissions. Am J Med Qual 2016 Dec;31(3):272-278. [doi: 10.1177/1062860614562627] [Medline: 25512952]
71. Crum AJ, Leibowitz KA, Verghese A. Making mindset matter. BMJ 2017 Feb 15;356:j674. [Medline: 28202443]
72. Dweck CS. Mindset: The New Psychology of Success. New York, New York: Random House Publishing Group; 2006.
73. Prensky M. From digital natives to digital wisdom: hopeful essays for 21st century learning. Thousand Oaks, CA: Corwin Press; 2012.
74. Hagen P, Collin P, Metcalf A, Nicholas M, Rahilly K, Swainston N. Pandora. Abbotsford, Vic: Young and Well Cooperative Research Centre; 2012. Participatory design of evidence-based online youth mental health promotion, intervention and treatment URL: http://pandora.nla.gov.au/pan/141862/20160405-1343/www.youngandwellcrc.org.au/wp-content/uploads/2014/03/Young_and_Well_CRC_IM_PD_Guide.pdf [accessed 2018-07-09] [WebCite Cache ID 70mKo3gUh]
75. Wong NT, Zimmerman MA, Parker EA. A typology of youth participation and empowerment for child and adolescent health promotion. Am J Community Psychol 2010 Sep;46(1-2):100-114. [doi: 10.1007/s10464-010-9330-0] [Medline: 20549334]
76. Mallan KM, Singh P, Giardina N. The challenges of participatory research with 'tech-savvy' youth. Journal of Youth Studies 2010 Feb 25;13(2):255-272. [doi: 10.1080/13676260903295059]
77. Robertson T, Simonsen J. Challenges and Opportunities in Contemporary Participatory Design. Design Issues 2012 Jul;28(3):3-9. [doi: 10.1162/DESI_a_00157]
78. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
79. Singh I, Wessely S. Childhood: a suitable case for treatment? Lancet Psychiatry 2015 Jul;2(7):661-666. [doi: 10.1016/S2215-0366(15)00106-6] [Medline: 26303563]
80. Hickie IB, Luscombe GM, Davenport TA, Burns JM, Highet NJ. Perspectives of young people on depression: awareness, experiences, attitudes and treatment preferences. Early Interv Psychiatry 2007 Nov;1(4):333-339. [doi: 10.1111/j.1751-7893.2007.00042.x] [Medline: 21352121]
81. Bradford S, Rickwood D. Adolescent's preferred modes of delivery for mental health services. Child Adolesc Ment Health 2012 Oct 11;19(1):39-45. [doi: 10.1111/camh.12002]
82. Barnett PA, Gotlib IH. Psychosocial functioning and depression: distinguishing among antecedents, concomitants, and consequences. Psychol Bull 1988 Jul;104(1):97-126. [Medline: 3043529]
83. Kawachi I, Berkman LF. Social ties and mental health. J Urban Health 2001 Sep;78(3):458-467 [FREE Full text] [doi: 10.1093/jurban/78.3.458] [Medline: 11564849]
84. Mermelstein R. Teen smoking cessation. Tob Control 2003 Jun;12 Suppl 1:i25-i34 [FREE Full text] [Medline: 12773783]
85. Fanshawe TR, Halliwell W, Lindson N, Aveyard P, Livingstone-Banks J, Hartmann-Boyce J. Tobacco cessation interventions for young people. Cochrane Database Syst Rev 2017 Dec 17;11:CD003289. [doi: 10.1002/14651858.CD003289.pub6] [Medline: 29148565]
86. Towns S, DiFranza JR, Jayasuriya G, Marshall T, Shah S. Smoking Cessation in Adolescents: targeted approaches that work. Paediatr Respir Rev 2017 Mar;22:11-22. [doi: 10.1016/j.prrv.2015.06.001] [Medline: 26187717]
87. Stockings E, Hall WD, Lynskey M, Morley KI, Reavley N, Strang J, et al. Prevention, early intervention, harm reduction, and treatment of substance use in young people. Lancet Psychiatry 2016 Mar;3(3):280-296. [doi: 10.1016/S2215-0366(16)00002-X] [Medline: 26905481]
88. Bootcamp Bootleg. Stanford, CA: Hasso Plattner Institute of Design at Stanford; 2010. URL: https://static1.squarespace.com/static/57c6b79629687fde090a0fdd/t/58890239db29d6cc6c3338f7/1485374014340/METHODCARDS-v3-slim.pdf [accessed 2018-07-09] [WebCite Cache ID 70mKYvyXY]
89. Balch GI, Tworek C, Barker DC, Sasso B, Mermelstein R, Giovino GA. Opportunities for youth smoking cessation: findings from a national focus group study. Nicotine Tob Res 2004 Feb;6(1):9-17. [doi: 10.1080/1462200310001650812] [Medline: 14982683]
90. Kealey KA, Ludman EJ, Mann SL, Marek PM, Phares MM, Riggs KR, et al. Overcoming barriers to recruitment and retention in adolescent smoking cessation. Nicotine Tob Res 2007 Feb;9(2):257-270. [doi: 10.1080/14622200601080315] [Medline: 17365757]

91. Leatherdale ST, McDonald PW. Youth smokers' beliefs about different cessation approaches: are we providing cessation interventions they never intend to use? Cancer Causes Control 2007 Sep;18(7):783-791. [doi: 10.1007/s10552-007-9022-8] [Medline: 17549592]
92. Allom V, Mullan B, Hagger M. Does inhibitory control training improve health behaviour? A meta-analysis. Health Psychol Rev 2016 Jun;10(2):168-186. [doi: 10.1080/17437199.2015.1051078] [Medline: 26058688]
93. Schoneveld EA, Lichtwarck-Aschoff A, Granic I. Preventing Childhood Anxiety Disorders: Is an Applied Game as Effective as a Cognitive Behavioral Therapy-Based Program? Prev Sci 2018 Feb;19(2):220-232 [FREE Full text] [doi: 10.1007/s11121-017-0843-8] [Medline: 28956222]
94. Schoneveld EA, Malmberg M, Lichtwarck-Aschoff A, Verheijen GP, Engels RC, Granic I. A neurofeedback video game (MindLight) to prevent anxiety in children: A randomized controlled trial. Computers in Human Behavior 2016 Oct;63:321-333. [doi: 10.1016/j.chb.2016.05.005]
95. Wijnhoven WA, Creemers DH, Engels RC, Granic I. The effect of the video game Mindlight on anxiety symptoms in children with an Autism Spectrum Disorder. BMC Psychiatry 2015 Jul 01;15:138 [FREE Full text] [doi: 10.1186/s12888-015-0522-x] [Medline: 26129831]
96. Tsui T. 2016. The efficacy of a novel video game intervention (MindLight) in reducing children's anxiety (Doctoral dissertation) URL: https://qspace.library.queensu.ca/handle/1974/14821 [accessed 2018-11-27] [WebCite Cache ID 74FIlTKxi]
97. Wols A, Lichtwarck-Aschoff A, Schoneveld EA, Granic I. In-Game Play Behaviours during an Applied Video Game for Anxiety Prevention Predict Successful Intervention Outcomes. J Psychopathol Behav Assess 2018;40(4):655-668 [FREE Full text] [doi: 10.1007/s10862-018-9684-4] [Medline: 30459485]
98. Novins DK, Green AE, Legha RK, Aarons GA. Dissemination and implementation of evidence-based practices for child and adolescent mental health: a systematic review. J Am Acad Child Adolesc Psychiatry 2013 Oct;52(10):1009-1025.e18 [FREE Full text] [doi: 10.1016/j.jaac.2013.07.012] [Medline: 24074468]
99. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of Intervention Principles: Evaluation Methods for Evolving Behavioral Intervention Technologies. J Med Internet Res 2015;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
100. Swendeman D, Comulada WS, Ramanathan N, Lazar M, Estrin D. Reliability and validity of daily self-monitoring by smartphone application for health-related quality-of-life, antiretroviral adherence, substance use, and sexual behaviors among people living with HIV. AIDS Behav 2015 Feb;19(2):330-340 [FREE Full text] [doi: 10.1007/s10461-014-0923-8] [Medline: 25331266]
101. van Reine PP. The culture of design thinking for innovation. J Innov Manag 2017;5:80.

Abbreviations

CBT: cognitive behavioral therapy
DMH: digital mental health interventions
DT: design thinking
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 10.07.18; peer-reviewed by B Auer, J Day, I Montagni; comments to author 03.10.18; accepted 22.10.18; published 14.01.19.

Please cite as:
Scholten H, Granic I
Use of the Principles of Design Thinking to Address Limitations of Digital Mental Health Interventions for Youth: Viewpoint
J Med Internet Res 2019;21(1):e11528
URL: https://www.jmir.org/2019/1/e11528/
doi:10.2196/11528
PMID:31344671

Viewpoint

Measuring the Implementation of Behavioral Intervention Technologies: Recharacterization of Established Outcomes

Eric DA Hermes[1,2], MD; Aaron R Lyon[3], PhD; Stephen M Schueller[4], PhD; Joseph E Glass[5], PhD, MSW

[1]Veterans Affairs Connecticut Healthcare System, West Haven, CT, United States

[2]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, United States

[3]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States

[4]Department of Psychological Science, University of California at Irvine, Irvine, CA, United States

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

Corresponding Author:
Eric DA Hermes, MD
Veterans Affairs Connecticut Healthcare System
950 Campbell Avenue
West Haven, CT, 06516
United States
Phone: 1 203 932 5711
Email: eric.hermes@yale.edu

Abstract

Behavioral intervention technologies (BITs) are websites, software, mobile apps, and sensors designed to help users address or change behaviors, cognitions, and emotional states. BITs have the potential to transform health care delivery, and early research has produced promising findings of efficacy. BITs also favor new models of health care delivery and provide novel data sources for measurement. However, there are few examples of successful BIT implementation and a lack of consensus on as well as inadequate descriptions of BIT implementation measurement. The aim of this viewpoint paper is to provide an overview and characterization of implementation outcomes for the study of BIT use in routine practice settings. Eight outcomes for the evaluation of implementation have been previously described: acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability. In a proposed recharacterization of these outcomes with respect to BIT implementation, definitions are clarified, expansions to the level of analysis are identified, and unique measurement characteristics are discussed. Differences between BIT development and implementation, an increased focus on consumer-level outcomes, the expansion of providers who support BIT use, and the blending of BITs with traditional health care services are specifically discussed. BITs have the potential to transform health care delivery. Realizing this potential, however, will hinge on high-quality research that consistently and accurately measures how well such technologies have been integrated into health services. This overview and characterization of implementation outcomes support BIT research by identifying and proposing solutions for key theoretical and practical measurement challenges.

(J Med Internet Res 2019;21(1):e11752) doi:10.2196/11752

KEYWORDS

mobile applications; behavior therapy; technology; internet; telemedicine; diffusion of innovation; translational medical research; outcome assessment (health care); review; implementation; behavioral intervention technology

Introduction

Behavioral Intervention Technology

A broad range of health information technologies are increasingly used in the delivery of health care to expand access, increase the effectiveness of care, and improve the productivity of health systems [1,2]. This article focuses on a subset of health information technology developed to intervene in a wide range of behavioral, psychosocial, or chronic health conditions, termed *behavioral health* conditions, by assisting the user to change behaviors, cognitions, and emotional states [3]. The term behavioral intervention technology (BIT) is used to refer to these interventions, although alternative terms such as eHealth, mobile health, *digital treatments*, and *internet interventions* are also used [4].

BITs are interventions delivered over computer software, internet websites, mobile apps, and wearable devices [2]. Such programs present material in varied formats, including audio,

video, text, or games. BITs may include symptom assessments, didactic lessons, passive sensing, and feedback systems that record and present a range of user activities and responses. BITs are used primarily by health care consumers but are also accessed by providers and others involved in the delivery of care. There are efficacious BITs for the treatment of most common behavioral health conditions, including depression, insomnia, substance use disorders, diabetes, and hypertension (Table 1) [5].

BITs represent an example of the move toward a more patient-centered health care system by empowering consumers to participate in their own care [1]. Some BITs provide novel data streams that providers can use to monitor patient outcomes, inform decision making, or improve care coordination. BITs may also increase access to and the convenience of care by reducing longstanding barriers such as travel, scheduling, and stigma while also potentially reducing health care costs [16-18]. Consequently, the capacity of BITs to support the so-called *triple aim* of health care reform (improved patient experience, population health, and costs) has spurred rapid growth in their development [19].

Nearly two decades of research on BITs has produced promising initial findings of efficacy; however, there are few examples of successful implementation and sustainment in routine practice settings [4,20]. For example, in a recent large pragmatic effectiveness trial in the United Kingdom, consumers used BITs for depression much less frequently than recommended, and the trial failed to demonstrate improvement in outcomes. However, a subsequent trial adding telephone support to BIT use did improve outcomes [21,22]. Given that use of such programs outside of research settings has been limited, there is a growing recognition that BITs will not revolutionize health care, without a better understanding of the factors associated with their implementation.

Objectives

In what has become a leading reference for the measurement of health service implementation, Proctor et al defined 8 outcomes for the measurement of implementation: acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability (or sustainment) [23]. The authors called for further work to conceptualize implementation outcomes across contexts and interventions. We heed this call and argue that traditional implementation outcomes must be recharacterized in light of the unique aspects of BITs and the models of health care delivery they favor. The objective of this article is to advance and clarify the diverse outcomes used to study the implementation of BITs.

This paper both relies on previous work in the area of BIT implementation and fills gaps in that work. Prior work has discussed the use of frameworks and theories to develop and implement health information technology [24,25]. This paper provides needed guidance on measuring the implementation of such technology. Moreover, several systematic reviews have evaluated how implementation outcomes have been measured for a subset of programs aimed at specific populations or a subset of implementation outcomes such as adoption and fidelity [4,26-28]. This paper provides a discussion of the full range of implementation outcomes with respect to a full range of BITs, addressing a host of behavioral health concerns. The goal of this article is to advance the evaluation of BIT implementation, with the hope of improving future implementation efforts and identifying key factors for successful implementation. We begin with a discussion of the types of services BITs provide to frame the argument. We then discuss each of Proctor et al's outcomes with regard to BIT implementation, followed by a discussion of measurement recommendations and future directions for BIT implementation work.

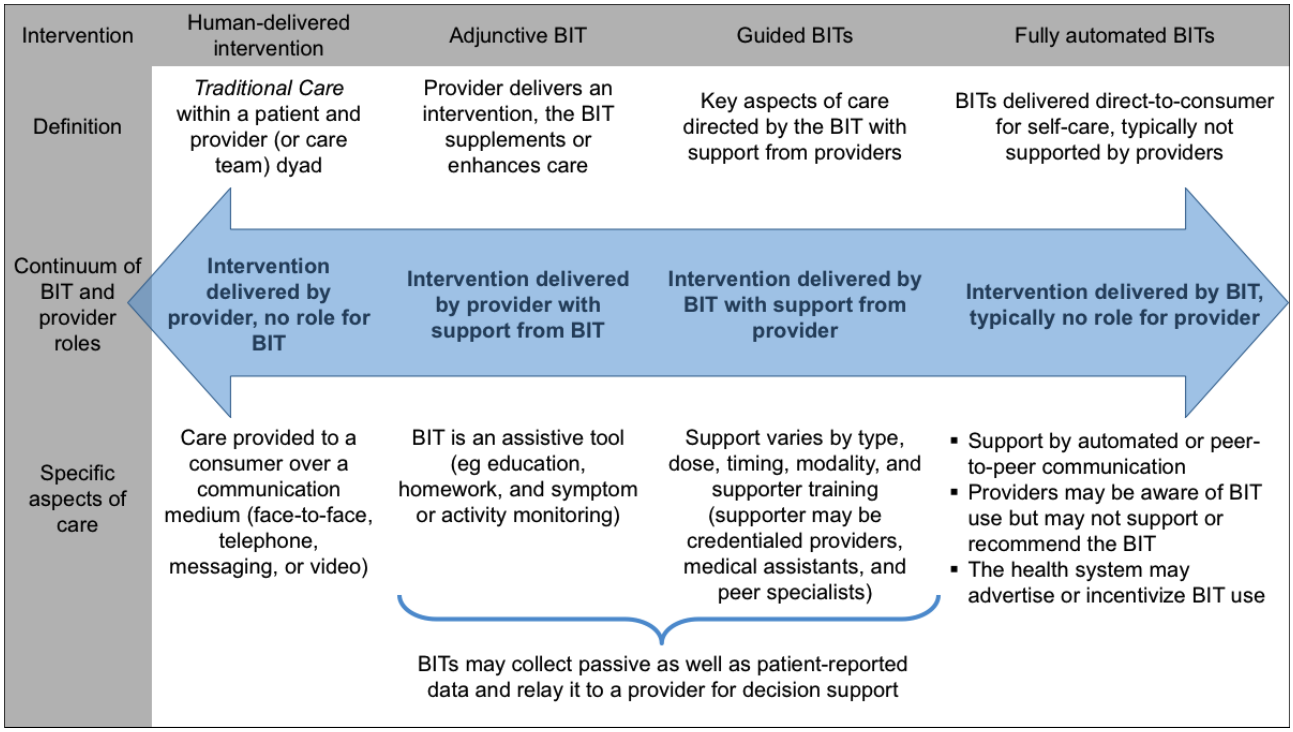
The Continuum of Health Services Supported by Behavioral Intervention Technology

The amount of clinical support provided to consumers using BITs has critical implications for the measurement of implementation outcomes. Some BITs are primarily consumer-facing products, designed to be used by consumers alone for care that is not directed by a health care provider (*self-care* or *fully automated* BITs, eg, MoodGYM [8,29]). Other BITs are intended to be used as a component of care that is delivered by a provider (*adjunctive* BITs; eg, the CBTi Coach mobile app [developed by US department of Veteran Affairs, Veterans Health Administration] as an adjunct to provider-led cognitive behavioral therapy for insomnia) [30]. As implementation research is focused on how practices are integrated into care settings, the level of provider integration with BIT use has critical implications for implementation outcome measurement. A description of the types of BITs and how they vary according to different levels of provider involvement is given in Figure 1, adapted from work by Muñoz (2017) [31].

Table 1. Examples of behavioral intervention technology programs.

Behavioral intervention technologies program	Program objective	Platform	Evidence
BlueStar or WellDoc	Diabetes management	Mobile	[6]
MoodGym	Depression	Web	[7,8]
Sleep Healthy Using the Internet (SHUTi)	Insomnia	Web	[9-11]
PTSD Coach and PE Coach	Posttraumatic stress disorder symptom tracking and treatment support	Mobile	[12-14]
reSET or Therapeutic Educational System	Substance use disorders	Mobile	[15]

Figure 1. Continuum of support for delivering behavioral intervention technologies (BITs).



Recharacterization of Implementation Outcomes for Behavioral Intervention Technology Implementation

Overview of Recharacterization

In a major advancement for implementation research, Proctor et al characterized 8 separate outcomes for the measurement of implementation [23]. In the following sections and Table 2, these implementation outcomes are recharacterized with respect to the unique aspects of BITs: the data streams produced by BITs, the continuum of support with which BITs are often employed (Figure 1), and the levels at which BIT implementation outcomes are analyzed (consumer, provider, administrator, and organization).

There are 3 important points to contextualize this recharacterization of BIT implementation outcomes. First, unless otherwise noted, we conceptualize implementation outcomes as applying to the clinical *intervention*—the BIT and the clinical activities intended to guide BIT use—rather than on the strategy or process used to implement the BIT. For example, the outcome of fidelity is limited to intervention fidelity (ie, the extent to which the BIT is delivered as intended) and not implementation fidelity (ie, the extent to which the strategy or process used to implement the BIT is followed as intended). Second, we consider BITs as interventions themselves rather than implementation strategies for other evidence-based practices (eg, an evidence-based self-care BIT for insomnia is an intervention and not an implementation strategy for cognitive behavioral therapy for insomnia). Although technology may intersect with the strategy used to implement an evidence-based practice, we do not consider the act of digitizing treatment to be an implementation strategy, as BITs require implementation strategies themselves to ensure successful implementation. Third, the construct of usability is pervasive in the technology design literature and partially overlaps with the implementation outcomes discussed. Building on the International Organization for Standardization's definition, we define BIT usability as the extent to which a BIT can be used to achieve the program's goals with accuracy, completeness, efficiency, and satisfaction in a specified context [32]. The concept of usability can be applied and assessed in all phases of BIT testing, from intervention development to preimplementation, implementation, and sustainment. BIT usability and its relationship with implementation outcomes is discussed more extensively below.

Implementation Outcomes

Acceptability

Acceptability is the extent to which an innovation is agreeable, palatable, or satisfactory to a stakeholder [23]. For BITs, acceptability can be evaluated among a variety of stakeholders (consumers, providers, administrators, and policy makers) and for a variety of BIT aspects, from intervention content to program appearance. A recent 2018 systematic review of implementation measurement for technology-based mental health interventions found that acceptability was the outcome most frequently measured, commonly via project-specific, nonvalidated measures [28].

Table 2. Characterization of behavioral intervention technology implementation outcomes.

Outcome and definition [23]	Level of analysis in BIT[a] studies	Measurement objective and process	Example of BIT outcome measurement
Acceptability			
• Perception among stakeholders that a given evidence-based practice is useful or satisfactory	• Individual provider, consumer, or administrator	• Objective: assessment of the extent to which BIT aligns with expectations of an agreeable user experience • Process: survey, interview, focus group, and direct observation usability testing	• Mares et al (2016): qualitative methods used to assess initial consumer and provider expectations [33] • Milward et al (2017): focus groups assessed the extent to which features were acceptable in terms of content, features, and design [34]
Adoption			
• Intention, decision, or initiation to use an evidence-based practice	• Individual provider, consumer, or administrator	• Objective: assessment of *actual system use* or *behavioral intention to use* • Process: passive data collection of BIT use [35]	• Gilbody et al (2016): to measure consumer-level adoption, log-in records were used to identify number of participants who accessed programs [22]
Appropriateness			
• Perceived fit, relevance, or compatibility of the evidence-based practice to a given context	• Individual provider, consumer, or administrator • Organization	• Objective: assessment of perceived BIT fit with the context • Process: survey, interview, focus group, direct observation usability testing, and workflow studies	• Lyon et al (2016): evaluated school-based practitioner workflows and current technology use practices to determine the appropriateness of a digital measurement feedback system and identified areas for BIT redesign [36]
Feasibility			
• Extent to which an evidence-based practice can be successfully used or conducted within a given context	• Individual provider, consumer, or administrator • Organization	• Objective: in vivo assessment of the extent to which a BIT can be used by consumers or providers in a specific setting • Process: Passive data collection of BIT use, survey, and structured observation studies	• Kumar et al (2018): using program use data collected via BIT, the feasibility of implementing a mobile app for consumers and a provider-facing dashboard was tested in 4 outpatient clinics [37]
Fidelity			
• Extent to which implementation results in an evidence-based practice being delivered as intended	• Individual consumer or provider • Organization	• Objective: measuring adherence, dose, or quality of BIT use with respect to the developer's intentions for use • Process: passive data collection of BIT use	• Calear et al (2013): reported high adherence associated with improved clinical outcomes in a hybrid implementation effectiveness study [7] • Sineath et al (2017): developed and tested a fidelity protocol for a diet and lifestyle monitoring BIT that involved coaching [38]
Implementation cost			
• Costs associated with implementing an evidence-based practice	• Organization	• Objective: intervention development costs, maintenance and versioning costs, implementation strategy costs, and operational costs • Process: cost analysis, interviews, and budgetary or administrative databases	• Quanbeck al (2018): measured implementation strategy costs for implementation coaching time and site visits needed to help 3 organizations integrate BIT into practice [39]

Outcome and definition [23]	Level of analysis in BIT[a] studies	Measurement objective and process	Example of BIT outcome measurement
Penetration			
• The integration of an evidence-based practice within a service setting (organization) and its subsystems	• Organization	• Objective: measuring the number of consumers or providers using BITs among those eligible or trained to engage in BIT • Process: passive data collection of BIT use and electronic health record data	• Titov et al (2015): measured the proportion of individuals in a defined consumer population completing lessons in 4 different BITs [40]
Sustainability			
• The extent to which a newly implemented evidence-based practice is maintained or institutionalized within a service setting's ongoing, stable operations	• Administrators • Organization	• Objectives: measuring ongoing BIT use, change in funding streams, saturation within the organization, and inclusion in routine reports • Process: passive data collection of BIT use, administrative or budgetary databases, oversight committee reports, and policy and training documents	• Carlfjord et al (2013): measured continued BIT delivery after active implementation [32] • Quanbeck et al (2018): measured whether or not the health system continued to offer BIT after research funding for an implementation trial ended [39]

[a]BIT: behavioral intervention technology.

The most well-known model of health information technology acceptability is the Technology Acceptance Model (TAM), in which the perceived usefulness and ease of use of the technology is assessed, along with a number of associated constructs such as prior experience, output quality, and social influence [35]. The concept of technology usability clearly overlaps with acceptability and is included as a major construct in the TAM and similar models. However, usability is a broader concept, encompassing additional issues such as likelihood of error and efficiency of use. In BIT research, usability is most often measured by self-report, which may be the component of usability that shares the most conceptual overlap with acceptability. Self-report measures of usability, such as the System Usability Scale, best track the notion of *perceived* usability (eg, Does a user believe that a BIT will be able to help them achieve a goal?) rather than *actual* usability, which involves observing whether those goals are achieved [41]. Validated measures such as the Attitudes Towards Psychological Online Interventions questionnaire also include assessments of acceptability [42]. Schroder et al (2017) used this measure to assess the acceptability of a BIT for depression among both consumers and providers in a randomized controlled trial [43].

Adoption

Proctor et al define adoption as the intention, decision, or initiation of use for an evidence-based practice, characterizing it at the level of the provider or organization [23]. Given the self-care nature of guided and fully automated BITs, this level of analysis is expanded to that of the consumer. The concept of adoption aligns with constructs of *actual system use* or b*ehavioral intention to use* in models such as TAM, for both consumers and providers [35].

The measurement of BIT adoption primarily relies on data gathered by the program, which is either passively collected (eg, log-in timestamp) or input by the consumer (eg, self-report symptom measure). However, direct observation and surveys of BIT use can occur. The amount and relative ease of data extraction from BITs can make intentions, decisions, or initiations more evident and accurate, without the need for labor-intensive data collection processes. In a large-scale naturalistic trial, Gilbody et al used log-in records to identify the number of study participants who accessed BITs for depression [22].

Appropriateness

Appropriateness is the perceived fit or compatibility of an innovation with a practice setting or context [23]. The key concept is the *perception* of fit, making the measurement of appropriateness more relevant to initial phases of implementation, but not strictly limited to these phases. Fit can be assessed at the organizational level (eg, alignment with workflow and policies) or the individual level (eg, alignment with providers' or consumers' attitudes, needs, and background). When evaluating implementation outcomes for BITs, we add administrators, given that BIT components such as dashboards can be accessed by administrators.

It is important to clearly specify which individuals or groups are intended to use a BIT and to elicit their needs when determining appropriateness. Usability evaluations can be employed to capture expectations, cognitions, and emotional responses of potential users [44]. Formative usability assessment seeks to assess the perceived contextual fit of new products with their destination context before actual implementation. Laboratory-based usability testing is particularly relevant to the initial assessment of appropriateness. Assessing the existing health information technology infrastructure, workflow, and use, as well as the potential interoperability of the BIT at a site, are key factors later in the process [45,46]. For instance, Lyon et al (2016) evaluated school-based practitioner workflows and current technology use practices to determine the

appropriateness of a digital measurement feedback system before its implementation [36].

Feasibility

Feasibility is the extent to which a new evidence-based practice can be successfully used or conducted in a setting. Feasibility differs from appropriateness as it is typically based on direct observation of stakeholder's experiences and practical concerns derived during or after implementation. Traditional conceptualizations of feasibility assessment include data collected from service providers and organizations and are rooted in the assumption that new innovations pass directly through providers to consumers [23]. For BITs, feasibility must be expanded to accommodate scenarios such as those involving fully automated BITs that may reach consumers with limited mediation from providers or the health system.

Standardized measures of innovation feasibility remain applicable to BITs [47]. In addition, although assessment of the acceptability and appropriateness of BITs are associated with formative usability testing, the assessment of BIT feasibility can leverage summative testing to evaluate how well full products meet their specified objectives in the actual context of use [48,49]. Such assessment might include evaluation of user awareness of key features of a BIT, actual compatibility with provider workflows, or user workarounds to account for unaddressed design issues. As feasibility is associated with users' actual experiences with an innovation in a specific context, in vivo usability testing is particularly relevant. Passive data collection opportunities surrounding feasibility may include frequency, time of day, or other contexts in which key BIT features are used. For instance, Lappalainen et al evaluated the feasibility of an adjunctive BIT for workplace stress reduction by determining how frequently users accessed an Web-based portal during the study [50].

Fidelity

Fidelity is the extent to which an intervention is used as intended and measured across several domains: protocol adherence, dose or amount delivered, and quality of delivery [23]. Protocol adherence to a BIT can be conceptualized as the functions of the program that were used, dose as the frequency of program use, and quality as whether the BIT was delivered correctly or for the intended purpose. The standards for these measurements should be based on the intentions of the BIT developers or empirical findings from efficacy testing. Therefore, fidelity to BITs suggests a match between the intended use of the program, termed *expected use* or *clinically meaningful use*, and its actual use by consumers [51]. Fidelity is traditionally measured at the provider level. However, the consumer level may be favored for BITs, especially for fully automated technologies.

BITs offer the opportunity to collect an extensive amount of passive data, which can be used to measure fidelity. Frequency metrics such as the number of log-ins, clicks, or task completions, as well as viewing times may best serve as measures of dose, whereas data representing the breadth of specific program functions used (eg, viewing of videos, completion of measures, and use of Web-based diaries) can serve as measures of adherence. Donkin et al provide a systematic review that includes measures of dose and adherence [27]. As for quality, BITs are a technology-driven platform, and they typically present information, direct activities, and make assessments in a uniform manner, which may provide better quality consistency compared with traditional human delivered care. Examples of measuring the quality dimension of fidelity include determining whether the BIT was used in a group of consumers with the appropriate condition, if the content of diary or self-report entries comply with the information requested, or whether passive sensors accurately assessed what was intended.

The measurement of certain aspects of fidelity, such as protocol adherence, may be less relevant for programs where consumers have more choice in when and how much they engage with the intervention. Determining *expected use* or *clinically meaningful use* with respect to BIT implementation is especially relevant and may be especially difficult for BITs developed as programs where users can openly navigate to browse information, activities, and assessments at their own discretion. This format differs from sequential navigation programs in which users are directed through a structured format to a specified endpoint [52]. Open navigation designs are increasingly common because of consumer desires for flexibility. This issue of clinically meaningful use and the level of fidelity associated with successful implementation is also relevant for programs used as an adjunct to provider-led care, where clinical outcomes may be associated more with provider activities than use of the BIT. A prime example is the Veterans Health Administration program *PTSD Coach*, a mobile phone BIT used in conjunction with provider-facilitated therapies for the treatment of posttraumatic stress disorder. The program has multiple functions, including the capability to monitor symptoms, learn relaxation techniques, and track appointments [12,13]. Lack of clarity around what constitutes *expected use* or *clinically meaningful use* might be one reason that research linking platform use and clinical outcomes has produced mixed findings [26,27].

Implementation Cost

Proctor et al identify 3 factors that drive costs in implementation research: the intervention, the setting of service delivery, and the strategy used to implement the intervention [23]. Intervention development and implementation may be blurred with respect to BITs, making these distinctions somewhat artificial. The decision to account for specific implementation costs in operational or implementation budgets will likely vary by organization but should be made clear in cost analyses.

BIT intervention costs include those associated with program development, testing, versioning, and maintenance. Program development and refinement also occurs with traditional face-to-face psychosocial interventions, but versioning and maintenance may be more applicable to BITs (eg, version 2.0 of a program) [53]. When developed outside the health system, intervention costs include the price paid by the health system for the BIT or to develop initial and subsequent versions of the BIT.

Costs associated with the context of service delivery will vary according to the implementation site's size, complexity, overhead, and how much adaptation of the site's current health information technology platform is needed. For instance,

Quanbeck et al measured system operating costs, which they saw as distinct from implementation costs, including staff time for introducing the BIT to patients and monitoring data produced by the BIT, technical support costs, and ongoing costs such as server hosting [39].

Implementation strategy costs are those incurred directly from the strategies used to employ the BIT, and vary depending on where the BIT lies on the continuum of support defined in Figure 1 [54]. Guided BITs may require more time to educate providers and consumers, facilitate engagement, and support use. In addition, resources such as advertising materials, educational materials, registries, clinical reminders, and decision support tools may be needed. Implementation strategies for fully automated BITs likely incur fewer costs given that providers do not require training and support to facilitate BIT use. In such cases, strategy costs may focus primarily on advertisement and other materials facilitating engagement.

Penetration

Penetration is the integration of a practice within a service setting and is measured within an organization. Some BITs allow for an expansion of service providers to include an array of providers who may support consumers during BIT use [23]. Other BITs may follow fully automated formats circumventing the involvement of providers and allowing for implementation outside of traditional health care systems. Thus, BITs may expand the breadth of providers for which penetration is measured and expand the reach of health care organizations beyond their *brick and mortar* facilities to draw in new consumer groups.

Penetration is primarily a summative implementation outcome traditionally measured as the number of providers who deliver a service among the total number of providers trained in or expected to deliver the service and/or the number of consumers engaging in a service among the number eligible to engage in a service within an organization [23]. Such traditional measurement applies well to adjunctive BITs. However, for guided BITs, penetration at the service provider level should be reconceptualized to include personnel specially trained to support BIT use, which comprises traditional health care delivery professionals as well as health technicians, administrative personnel, and peers [29]. The decision to include such adjunct providers in the larger denominator of providers, or measure penetration separately according to provider role is up to the researcher. We note that in such cases, the number of adjunct providers trained to support BIT use may be small, especially if they are centralized to provide remote support across a health care organization. Consumer-level penetration or fidelity to BIT delivery may be a more impactful measure in such cases.

Some fully automated BITs may be implemented outside the traditional health system contexts at the consumer level, either *adjacent* to health care systems to specifically targeted populations or disseminated worldwide as in Massive Open Online Interventions (MOOIs), where specific targeting strategies are not used [55]. In the case of targeted populations, the denominator for penetration must be selected from those for which the intervention is clinically appropriate and to which the outreach efforts for the intervention were aimed. Measuring MOOI penetration may be difficult, as the denominator cannot be precisely calculated, and estimated penetration, based on the proportion reached given the prevalence of a targeted condition, may be more appropriate.

Sustainability

Sustainability is the degree to which an implemented treatment is maintained, institutionalized, or integrated within a service setting. For BITs, sustainment is characterized at the level of the organization or setting. Proctor et al distinguished 3 constructs for the measurement of sustainability: passage, niche saturation, and inclusion in cycles or routines [23].

The implementation of BITs in US health care systems has primarily taken place as effectiveness trials and pilot programs. Thus, sustainment is the least well-documented BIT implementation outcome [4]. As BIT use matures, passage from research and development funding streams to permanent organizational funding will be important to measure. In addition, the integration of BITs with pre-existing organizational health information technology such as electronic health records or other patient or provider dashboards will be an important measure of niche saturation. Similarly, the addition of metrics on BIT use and clinical outcomes in routine administrative reports will be an important measure of inclusion in cycles or routines.

With respect to fully automated BITs implemented outside of traditional health care settings, the analysis of sustainment beyond simple continued adoption may be difficult because of the novel organizational cultures in these contexts. As a hypothetical example, analysis of the sustainment of a BIT aimed at reducing alcohol intake among veterans and implemented through social media by a veteran service organization may rely on analysis of the organization's continued budget for advertising and program maintenance as a gauge of sustainment, in addition to routine adoption or fidelity measures.

An additional unique aspect of BITs is the concept of intervention versioning, whereby there is a continuous evolution of an intervention through frequent program updates [56]. Some changes are small (ie, bug fixes or cosmetic improvements), whereas others may include substantial changes to program content and function. Measurement of implementation and specifically sustainment of BIT use must account for such changes. At a minimum, the timing and content of changes should be tracked and associated with any changes in BIT use and clinical outcome.

Recommendations for Behavioral Intervention Technology Implementation Measurement

A key factor in BIT implementation measurement involves the level of provider integration. Fully automated BITs are platforms for patient-centered self-care, the implementation of which may circumvent traditional health care organizations. As such, implementation measurement for fully automated BITs should

emphasize the consumer level for outcomes such as adoption, and the organizational level, which may include nontraditional health care delivery organizations, for outcomes such as cost and sustainment. In guided BITs, care is directed by the BIT and support is delivered by a provider. Implementation outcomes are still primarily measured at the consumer level. However, the measurement of provider or coach activities is also important, especially with respect to fidelity, implementation cost, and sustainment. In adjunctive BITs, BIT use constitutes an evidence-based practice that supports aspects of provider-led care. The traditional provider-level focus of implementation outcome measurement still applies. However, the BIT may allow unique ways to measure such outcomes. For example, the use of an adjunctive BIT might increase fidelity to an evidence-based practice as well as provide objective information on that fidelity through passive data collection. On the basis of the above discussion of BIT implementation outcomes, we have identified additional recommendations for the measurement of BIT implementation in Textbox 1 and in the proposed agenda for BIT implementation research below.

Textbox 1. Implementation outcome and recommendations for measuring behavioral intervention technology (BIT) implementation.

Implementation outcome and recommendations

Acceptability

- Articulate measurement distinctions between acceptability and usability.
- Use validated measures of acceptability.

Adoption

- Expand measurement to the consumer level, in addition to providers in the case of guided and adjunctive BITs.
- Define and measure adoption using BIT data streams before and during the implementation effort.

Appropriateness

- Evaluate usability or workflow to assess appropriateness.
- Assess user perceptions of appropriateness in the initial phases of implementation.

Feasibility

- Measure using data acquired through BIT data streams at the consumer and provider levels.
- Define and measure feasibility before and during the implementation effort.

Fidelity

- Consider measurement approaches that include dose, adherence, or quality.
- Measure at the consumer level to ensure the course of treatment (BIT usage) is followed as intended.
- Measure at the provider level to ensure the adequacy of BIT guidance, support, or coaching.
- Base standards for fidelity on intentions of developers or empirical findings from BIT testing.
- Develop clear fidelity standards for openly navigated programs.

Implementation cost

- Assess BIT costs (development, testing, and versioning); context costs; and implementation strategy costs.

Penetration

- Measure primarily among consumers and providers.
- Include providers who support BIT use in guided BITs.
- Include groups targeted for BIT use in the denominator.

Sustainability

- Measure passage to permanent funding, integration with pre-existing technology, and inclusion in routine reports, in addition to continued BIT use beyond active implementation.
- Attend explicitly to BIT versioning and updates.

Agenda for Behavioral Intervention Technology Implementation Research

Agenda Overview

BITs are novel platforms for behavioral health interventions that promote patient-centered care by blending technology-supported self-care with traditional health services led by providers. In spite of the potential benefits of BITs and the well-documented efficacy of many BITs, their successful implementation has faced considerable challenges [22,28]. To help advance the study of factors associated with successful BIT implementation, there should first be consensus on the measurement of BIT implementation outcomes. To this end, we have selected several characteristics of BIT implementation on which future research should be based.

Focus on Implementation Outcomes at the Consumer Level

BITs are 1 aspect of the current transition in health care delivery from services focused on traditional provider-centered face-to-face care in *brick-and-mortar* facilities to patient-centered services that include self-care [1,57]. This new form of care may be better described as *brick-and-click*, as it incorporates the traditional provider-led care in health care facilities, with the *clicks* associated with consumer-led health information technology use. Consequently, the implementation of patient-centered digital interventions will require greater focus on and more precise measurement of consumer-level implementation outcomes while continuing to assess the provider- and organizational-level implementation outcomes that have been the traditional focus.

For example, there is a need for more precise characterization of the different aspects of BIT adoption and fidelity. BIT developers and implementation researchers must clearly define what consumer-level adoption means for a specific BIT or type of BIT. Is the most appropriate measure of consumer adoption an initial program log-in, a specific number of clicks or pages visited, or a threshold of elapsed engagement time? Alternatively, stages, milestones, or levels of adoption can be quantified based on the engagement in or completion of certain activities, similar to the stages of implementation completion [58].

Fidelity must also be clearly defined for a given BIT, especially for programs that allow users to access a range of resources in a self-directed manner. The level of use that constitutes adequate fidelity should ideally be specified before program implementation. For example, fidelity to a BIT modeled after a manualized cognitive behavioral therapy for insomnia may be relatively straightforward: completing the program's 6 sessions, associated homework, and daily sleep diary entries [11]. However, determining adequate fidelity to a BIT that does not dictate how or how much it should be used, such as the *Vets Prevail* program where consumers access a range of activities for depression and anxiety symptoms in a nonlinear fashion, may be more difficult [59]. Ultimately, the association between fidelity and clinical outcomes must be measured [27].

Behavioral Intervention Technology Usability Assessment and Implementation Outcomes

There is a conceptual and functional overlap between BIT usability assessment and implementation outcomes. We see usability as a broad concept centered primarily in program design and development phases but with clear elements overlapping with the early stages of implementation. We hypothesize predictive associations between (1) usability assessment in the development of BITs; (2) implementation outcomes of acceptability, appropriateness, and feasibility; and (3) subsequent downstream implementation outcomes of BIT adoption, penetration, and sustainment [60]. These hypotheses need empirical investigation. Moreover, the overlap between the conceptual and practical aspects of usability and implementation outcomes highlights potential for ambiguity among the implementation outcomes described. As others have proposed, there may be theoretical divisions between these outcomes but little empirical evidence or practical means to measure distinctions between them [61]. For instance, some have determined feasibility by measuring the frequency of logging in to a BIT portal, whereas others may consider this a measure of consumer adoption [50].

Rapid Behavioral Intervention Technology Development

Similarly, the transition from intervention development and testing to implementation is accelerated and more iterative for BITs compared with traditional services [56]. In some ways, the technology aspects of BITs, including program updates and versioning may allow for more consistent delivery of services, but rapid development and versioning poses problems for evaluation and implementation. For example, the relative ease of deployment may have led to the attempted implementation of some BIT programs or updates without evidence of usability or efficacy [62,63]. Such situations may lead to innovation fatigue among consumers and providers, potentially undermining successful sustainment [64]. The timing of implementation with respect to evidence of innovation usability and efficacy is a topic of interest in implementation science as a whole and deserves more investigation with respect to BITs.

Expansion of Providers Who Support Behavioral Intervention Technology use

The implementation of BITs will also broaden the definition of care providers to include individuals who support care through BITs, such as health technicians, administrative personnel, peers, and other paraprofessionals [65,29]. Implementation outcome measurement, especially fidelity, must accommodate the activity and perspectives of these individuals. The measurement of this activity may be nuanced or difficult, as BITs may be implemented outside of traditional health care contexts, the consumer and provider may be geographically separated, the provider may not be employed by the health system, or communication maybe infrequent or through nontraditional means such as short message service text messages. Implementation researchers must be aware of these contingencies and clearly define the roles of different providers as well as metrics to capture their actions.

Combining Behavioral Intervention Technologies With Traditional Health Care Services

As described in Figure 1, BITs integrate their treatment function across a continuum of human involvement, from provider-led services (adjunctive BITs) to consumer-facing self-care services (fully automated BITs). Future research should better characterize the breadth of this continuum as well as ramifications for BIT implementation outcome measurement, implementation strategy development, and the empirical determination of clinically meaningful use. For example, implementing adjunctive BITs requires acceptance and adoption of technology by the consumer, provider, and organization, with relatively little change in individual roles. The implementation of such programs may not be unlike the implementation of similar adjunctive services such as blood pressure or glucose monitoring devices. However, BITs that support a high degree of self-care, such as guided BITs, will likely require a greater degree of change to the basic function and workflow of health care delivery organizations, requiring more robust implementation strategies. Such relationships should be investigated in future BIT implementation research.

Stepped care models, where low-intensity treatment is initiated first followed by elevations in treatment intensity as needed, are an intuitive approach to integrating BITs in health care systems, where BITs serve as the initial stage before more intensive face-to-face treatment [66]. It is also possible to integrate stepped care across the continuum of BIT support. For instance, fully automated BITs could serve as a low-intensity initial step, with patients being stepped up to more intensive coaching or other forms of support as needed. Such adaptive implementation strategies may be particularly suited to BITs and testing through factorial or sequential multiple assignment randomized trial designs [67]. In such trials, BIT Implementation researchers should clearly define the adoption, fidelity, or clinical outcomes that necessitate transitions.

Downstream Effects on Health Service Utilization

One implicit assumption of those advocating the use of self-care BITs through stepped care models is that the use of such services represents an alternative pathway or entry point to care [1]. Although we support this position, we believe analyzing the impact of BIT use in addressing unmet need and subsequent or alternative care utilization is an important area of further research. For instance, there may be sunk or opportunity costs of time for consumers who initially engage in a BIT and have difficulty, subsequently dropping out [68]. Such a failure may engender an attitude of therapeutic nihilism on the part of the consumer or even provider, who may lose motivation and limit further engagement in behavioral treatment [69]. However, such potential costs are largely theoretical and deserve further research as initial evidence has shown that engagement in face-to-face care may increase for consumers who also receive a BIT [15,70].

Conclusions

BITs are part of the current patient-centered transformation in health care delivery but initial implementation attempts have had varied outcomes, and more rigorous measurement of BIT implementation is needed to advance the understanding of factors related to successful implementation. Such implementation outcomes must be recharacterized to account for the unique aspects of BITs, as Proctor et al suggested in their original work [23]. Through this recharacterization, we have identified areas where the field of BIT implementation science can be advanced further. More work, interaction, and debate in these areas will allow for the exploration of the empirical boundaries primarily between BIT's development-related outcomes such as usability and implementation- related outcomes such as acceptability, feasibility, and appropriateness. Further studies can differentiate the conceptual and practical differences in these constructs as well as quantify hypothesized associations between them. Future work should also more clearly and empirically define the outcomes of adoption and fidelity with respect to BITs. Finally, work aimed at developing, defining, and testing strategies for implementing different BITs in varying contexts is needed. Our recharacterization of implementation outcomes with respect to BITs intends to potentiate this work.

Acknowledgments

EDAH is supported by a Veterans Health Affairs, Health Services Research and Development (HSR&D) Career Development Award, the Veterans Affairs (VA) New England Mental Illness Research, Education, and Clinical Center, as well as The VA Pain, Research, Informatics, Medical comorbidities, and Education Center. SMS is supported by a Career Development Award from the National Institute of Mental Health. JEG is supported by the National Institutes of Health Extramural Loan Repayment Program for Health Disparities Research (L60 MD009373) and a Mentored Research Scientist Development Award (K01 AA023859). The preparation of this paper was supported in part by the Implementation Research Institute at the George Warren Brown School of Social Work, Washington University in St Louis, through an award from the National Institute of Mental Health (5R25MH08091607) and the Department of Veterans Affairs, HSR&D Service, and Quality Enhancement Research Initiative. The funding sources had no role in the design, analysis, or interpretation of data or in the preparation of the report or decision to publish, nor had any role in the views expressed. All authors have reviewed and approved this manuscript for publication.

Conflicts of Interest

None declared.

References

1. Fortney JC, Burgess JF, Bosworth HB, Booth BM, Kaboli PJ. A re-conceptualization of access for 21st century healthcare. J Gen Intern Med 2011 Nov;26(Suppl 2):639-647 [FREE Full text] [doi: 10.1007/s11606-011-1806-6] [Medline: 21989616]
2. Marsch LA, Lord SE, Dallery J, editors. Behavioral healthcare and technology: using science-based innovations to transform practice. Oxford, UK: Oxford University Press; 2014:368.
3. England MJ, Butler AS, Gonzalez ML, editors. Psychosocial interventions for mental and substance use disorders: A framework for establishing evidence-based standards. Washington, DC: National Academy Press; 2015:209.
4. 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 Jun 6;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
5. Sander L, Rausch L, Baumeister H. Effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. J Med Internet Res Ment Health 2016;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
6. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc™ mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 May 12;10(3):160-168 [FREE Full text] [doi: 10.1089/dia.2008.0283]
7. Calear AL, Christensen H, Mackinnon A, Griffiths KM. Adherence to the MoodGYM program: outcomes and predictors for an adolescent school-based population. J Affect Disord 2013 May;147(1-3):338-344. [doi: 10.1016/j.jad.2012.11.036] [Medline: 23245469]
8. Schneider J, Sarrami Foroushani P, Grime P, Thornicroft G. Acceptability of online self-help to people with depression: users' views of MoodGYM versus informational websites. J Med Internet Res 2014 Mar 28;16(3):e90 [FREE Full text] [doi: 10.2196/jmir.2871] [Medline: 24681717]
9. Christensen H, Batterham PJ, Gosling JA, Ritterband LM, Griffiths KM, Thorndike FP, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. Lancet Psychiatry 2016 Apr;3(4):331-341. [doi: 10.1016/S2215-0366(15)00536-2] [Medline: 26827250]
10. Ritterband LM, Thorndike FP, Gonder-Frederick LA, Magee JC, Bailey ET, Saylor DK, et al. Efficacy of an internet-based behavioral intervention for adults with insomnia. Arch Gen Psychiatry 2009 Jul;66(7):692-698 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2009.66] [Medline: 19581560]
11. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a web-based cognitive behavior therapy for insomnia intervention with 1-year follow-up: a randomized clinical trial. J Am Med Assoc Psychiatry 2017 Jan 01;74(1):68-75. [doi: 10.1001/jamapsychiatry.2016.3249] [Medline: 27902836]
12. 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]
13. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016;38:94-98. [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]
14. Reger GM, Hoffman J, Riggs D, Rothbaum BO, Ruzek J, Holloway KM, et al. The "PE coach" smartphone application: an innovative approach to improving implementation, fidelity, and homework adherence during prolonged exposure. Psychol Serv 2013 Aug;10(3):342-349. [doi: 10.1037/a0032774] [Medline: 23937084]
15. Campbell AN, Nunes EV, Matthews AG, Stitzer M, Miele GM, Polsky D, et al. Internet-delivered treatment for substance abuse: a multisite randomized controlled trial. Am J Psychiatry 2014 Jun;171(6):683-690 [FREE Full text] [doi: 10.1176/appi.ajp.2014.13081055] [Medline: 24700332]
16. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
17. Murphy SM, Campbell AN, 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 1;161:119-126 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.01.021] [Medline: 26880594]
18. 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] [Medline: 26813512]
19. Sheikh A, Sood HS, Bates DW. Leveraging health information technology to achieve the "triple aim" of healthcare reform. J Am Med Inform Assoc 2015 Jul;22(4):849-856 [FREE Full text] [doi: 10.1093/jamia/ocv022] [Medline: 25882032]
20. 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] [Medline: 28341617]
21. 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] [Medline: 28254959]

22. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, et al. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. Br Med J 2016 Jan 12;352:i195 [FREE Full text] [Medline: 26759375]
23. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
24. Glasgow RE, Dickinson P, Fisher L, Christiansen S, Toobert DJ, Bender BG, et al. Use of RE-AIM to develop a multi-media facilitation tool for the patient-centered medical home. Implement Sci 2011 Oct 21;6:118 [FREE Full text] [doi: 10.1186/1748-5908-6-118] [Medline: 22017791]
25. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. Int J Med Inform 2014 Jul;83(7):e1-11. [doi: 10.1016/j.ijmedinf.2013.07.002] [Medline: 23910896]
26. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
27. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
28. Wozney L, McGrath PJ, Gehring ND, Bennett K, Huguet A, Hartling L, et al. eMental healthcare technologies for anxiety and depression in childhood and adolescence: systematic review of studies reporting implementation outcomes. J Med Internet Res Ment Health 2018 Jun 26;5(2):e48 [FREE Full text] [doi: 10.2196/mental.9655] [Medline: 29945858]
29. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions? A systematic review. Internet Interv 2014 Jan;1(4):205-215 [FREE Full text] [doi: 10.1016/j.invent.2014.08.003]
30. Kuhn E, Weiss BJ, Taylor KL, Hoffman JE, Ramsey KM, Manber R, et al. CBT-I coach: a description and clinician perceptions of a mobile app for cognitive behavioral therapy for insomnia. J Clin Sleep Med 2016;12(4):597-606. [doi: 10.5664/jcsm.5700] [Medline: 26888586]
31. Muñoz RF. The efficiency model of support and the creation of digital apothecaries. Clin Psychol Sci Pract 2016 Nov 17;24(1):46-49 [FREE Full text] [doi: 10.1111/cpsp.12174]
32. International Organization for Standardization. 2010 Mar. Ergonomics of human-system interaction-part 210: Human-centered design for interactive systems URL: https://www.iso.org/standard/52075.html [accessed 2018-12-10] [WebCite Cache ID 74YfDMJ3L]
33. Mares ML, 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] [Medline: 27687632]
34. Milward J, Deluca P, Drummond C, Watson R, Dunne J, Kimergård A. Usability testing of the BRANCH smartphone app designed to reduce harmful drinking in young adults. J Med Internet Res Mhealth Uhealth 2017 Aug 8;5(8):e109 [FREE Full text] [doi: 10.2196/mhealth.7836] [Medline: 28790022]
35. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep;27(3):425-478. [doi: 10.2307/30036540]
36. Lyon AR, Ludwig K, Wasse JK, Bergstrom A, Hendrix E, McCauley E. Determinants and functions of standardized assessment use among school mental health clinicians: a mixed methods evaluation. Adm Policy Ment Health 2016 Jan;43(1):122-134 [FREE Full text] [doi: 10.1007/s10488-015-0626-0] [Medline: 25875325]
37. Kumar D, Tully LM, Iosif AM, Zakskorn LN, Nye KE, Zia A, et al. A mobile health platform for clinical monitoring in early psychosis: implementation in community-based outpatient early psychosis care. J Med Internet Res Ment Health 2018 Feb 27;5(1):e15 [FREE Full text] [doi: 10.2196/mental.8551] [Medline: 29487044]
38. Sineath A, Lambert L, Verga C, Wagstaff M, Wingo BC. Monitoring intervention fidelity of a lifestyle behavioral intervention delivered through telehealth. Mhealth 2017;3:35 [FREE Full text] [doi: 10.21037/mhealth.2017.07.04] [Medline: 28894745]
39. 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] [Medline: 29382624]
40. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. MindSpot clinic: an accessible, efficient, and effective online treatment service for anxiety and depression. Psychiatr Serv 2015 Oct;66(10):1043-1050. [doi: 10.1176/appi.ps.201400477] [Medline: 26130001]
41. Brooke J. SUS-A quick and dirty usability scale. In: Usability evaluation in industry. Boca Raton: CRC Press; Jun 1996.
42. Schröder J, Sautier L, Kriston L, Berger T, Meyer B, Späth C, et al. Development of a questionnaire measuring attitudes towards psychological online interventions-the APOI. J Affect Disord 2015 Nov 15;187:136-141. [doi: 10.1016/j.jad.2015.08.044] [Medline: 26331687]
43. Schröder J, Berger T, Meyer B, Lutz W, Hautzinger M, Späth C, et al. Attitudes towards internet interventions among psychotherapists and individuals with mild to moderate depression symptoms. Cogn Ther Res 2017 Oct;41(5):745-756. [doi: 10.1007/s10608-017-9850-0]

44. Grudin J, Pruitt J. Personas, participatory design and product development: An Infrastructure for Engagement. In: Proceedings of Participation and Design Conference (PDC2002). 2002 Presented at: Participation and Design Conference (PDC2002); June 23-25, 2002; Sweden p. 144-161.
45. Lyon AR, Wasse JK, Ludwig K, Zachry M, Bruns EJ, Unützer J, et al. The contextualized technology adaptation process (CTAP): optimizing health information technology to improve mental health systems. Adm Policy Ment Health 2016 May;43(3):394-409. [doi: 10.1007/s10488-015-0637-x] [Medline: 25677251]
46. Zheng K, Haftel HM, Hirschl RB, O'Reilly M, Hanauer DA. Quantifying the impact of health IT implementations on clinical workflow: a new methodological perspective. J Am Med Inform Assoc 2010;17(4):454-461 [FREE Full text] [doi: 10.1136/jamia.2010.004440] [Medline: 20595314]
47. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017 Dec 29;12(1):108 [FREE Full text] [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
48. Rubin J, Chisnell D, Spool J. Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. New York, NY: Wiley, John & Sons; 2008:384.
49. Zhou X, Qiu Y, Hua G, Wang H, Ruan X. A feasible approach to the integration of CAD and CAPP. Comput Aided Des 2007 Apr;39(4):324-338 [FREE Full text] [doi: 10.1016/j.cad.2007.01.005]
50. Lappalainen P, Kaipainen K, Lappalainen R, Hoffrén H, Myllymäki T, Kinnunen ML, et al. Feasibility of a personal health technology-based psychological intervention for men with stress and mood problems: randomized controlled pilot trial. J Med Inernet Res Res Protoc 2013 Jan 9;2(1):e1 [FREE Full text] [doi: 10.2196/resprot.2389] [Medline: 23611946]
51. Morrison C, Doherty G. Analyzing engagement in a web-based intervention platform through visualizing log-data. J Med Internet Res 2014;16(11):e252 [FREE Full text] [doi: 10.2196/jmir.3575] [Medline: 25406097]
52. 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] [Medline: 16403040]
53. 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]
54. 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]
55. Muñoz RF, Bunge EL, Chen K, Schueller SM, Bravin JI, Shaughnessy EA, et al. Massive open online interventions: a novel model for delivering behavioral-health services worldwide. Clin Psychol Sci 2016;4(2):194-205 [FREE Full text] [doi: 10.1177/2167702615583840]
56. Mohr DC, Cheung K, Schueller SM, Hendricks Brown C, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
57. Tang PC, Smith MD. Democratization of health care. J Am Me Assoc 2016 Oct 25;316(16):1663-1664. [doi: 10.1001/jama.2016.14060] [Medline: 27669345]
58. Chamberlain P, Brown CH, Saldana L. Observational measure of implementation progress in community based settings: the stages of implementation completion (SIC). Implement Sci 2011 Oct 6;6:116 [FREE Full text] [doi: 10.1186/1748-5908-6-116] [Medline: 21974914]
59. Hobfoll SE, Blais RK, Stevens NR, Walt L, Gengler R. Vets prevail online intervention reduces PTSD and depression in veterans with mild-to-moderate symptoms. J Consult Clin Psychol 2016 Jan;84(1):31-42. [doi: 10.1037/ccp0000041] [Medline: 26322788]
60. Lyon, Bruns E. User-centered redesign of evidence-based psychosocial interventions to enhance implementation-hospitable soil or better seeds? JAMA Psychiatry 2018 Nov 14:- Epub ahead of print(forthcoming). [doi: 10.1001/jamapsychiatry.2018.3060] [Medline: 30427985]
61. Lewis CC, Fischer S, Weiner BJ, Stanick C, Kim M, Martinez RG. Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. Implement Sci 2015 Nov 4;10(1):155 [FREE Full text] [doi: 10.1186/s13012-015-0342-x] [Medline: 26537706]
62. 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] [Medline: 24240579]
63. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
64. Chung GH, Choi JN, Du J. Tired of innovations? Learned helplessness and fatigue in the context of continuous streams of innovation implementation. J Organ Behav 2017 Sep;38(7):1130-1148 [FREE Full text] [doi: 10.1002/job.2191]

65. Schueller SM, Tomasino KN, Mohr DC. Integrating human support into behavioral intervention technologies: the efficiency model of support. Clin Psychol 2017;24(1):27-45 [FREE Full text] [doi: 10.1111/cpsp.12173]
66. Katon W, Von Korff M, Lin E, Simon G, Walker E, Unützer J, et al. Stepped collaborative care for primary care patients with persistent symptoms of depression: a randomized trial. Arch Gen Psychiatry 1999 Dec;56(12):1109-1115. [Medline: 10591288]
67. Almirall D, Compton SN, Gunlicks-Stoessel M, Duan N, Murphy SA. Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. Stat Med 2012 Jul 30;31(17):1887-1902 [FREE Full text] [doi: 10.1002/sim.4512] [Medline: 22438190]
68. Rozental A, Boettcher J, Andersson G, Schmidt B, Carlbring P. Negative effects of internet interventions: a qualitative content analysis of patients' experiences with treatments delivered online. Cogn Behav Ther 2015;44(3):223-236. [doi: 10.1080/16506073.2015.1008033] [Medline: 25705924]
69. Mamede S, Schmidt HG. The twin traps of overtreatment and therapeutic nihilism in clinical practice. Med Educ 2014 Jan;48(1):34-43. [doi: 10.1111/medu.12264] [Medline: 24330115]
70. Glass JE, McKay JR, Gustafson DH, Kornfield R, Rathouz PJ, McTavish FM, et al. Treatment seeking as a mechanism of change in a randomized controlled trial of a mobile health intervention to support recovery from alcohol use disorders. J Subst Abuse Treat 2017 Jun;77:57-66. [doi: 10.1016/j.jsat.2017.03.011] [Medline: 28476273]

Abbreviations

BIT: behavioral intervention technology
MOOI: massive open online intervention
TAM: technology acceptance model

Edited by G Eysenbach; submitted 03.08.18; peer-reviewed by P McGee-Vincent, T Fazzino; comments to author 03.10.18; revised version received 18.10.18; accepted 20.10.18; published 25.01.19.

Please cite as:
Hermes EDA, Lyon AR, Schueller SM, Glass JE
Measuring the Implementation of Behavioral Intervention Technologies: Recharacterization of Established Outcomes
J Med Internet Res 2019;21(1):e11752
URL: http://www.jmir.org/2019/1/e11752/
doi:10.2196/11752
PMID:30681966

Tutorial

Human-Centered Design of Video-Based Health Education: An Iterative, Collaborative, Community-Based Approach

Maya Adam[1], MD; Shannon A McMahon[2,3], MHS, PhD; Charles Prober[1], MD; Till Bärnighausen[2,4,5], MD, ScD

[1]Stanford Center for Health Education, Stanford School of Medicine, Stanford University, Stanford, CA, United States

[2]Institute of Global Health, Heidelberg University, Heidelberg, Germany

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

[4]Harvard T.H. Chan School of Public Health, Department of Global Health and Population, Harvard University, Boston, MA, United States

[5]Africa Health Research Institute, Wellcome Trust, Durban, South Africa

Corresponding Author:
Maya Adam, MD
Stanford Center for Health Education
Stanford School of Medicine
Stanford University
408 Panama Mall
Stanford, CA, 94305
United States
Phone: 1 650 839 3600
Email: madam@stanford.edu

Abstract

Drawing on 5 years of experience designing, producing, and disseminating video health education programs globally, we outline the process of creating accessible, engaging, and relevant video health education content using a community-based, human-centered design approach. We show that this approach can yield a new generation of interventions, which are better aligned with the needs and contexts of target communities. The participation of target communities and local stakeholders in the content production and design process fosters ownership of the content and increases the likelihood that the resulting intervention will resonate within its intended primary audience and be disseminated broadly. Ease of future adaptation for additional global audiences and modification of the content for multiple dissemination pathways are important early considerations to ensure scalability and long-term impact of the intervention. Recent advances in mobile technology can facilitate the dissemination of accessible, engaging health education at scale, thereby enhancing the potential impact of video-based educational tools. Accessible and engaging health education is a cornerstone of health behavior change. Especially in low- and middle-income countries, increasing access to effective health education can contribute to improved health outcomes. Prior research has identified several characteristics of effective health education interventions. These include the integration of pictures, narratives, and entertainment-education, in which the health messages that make up the educational content are embedded. However, the effectiveness and long-term impact of health messages ultimately depend on how well the end users can identify with the content that is presented. This identification, in turn, is a function of how well the messages correspond to user needs and wants and how this correspondence is communicated through the design characteristics of the health education intervention.

(J Med Internet Res 2019;21(1):e12128) doi:10.2196/12128

KEYWORDS

human-centered design; health promotion; health behavior; health knowledge, attitudes, practice; community health workers; telemedicine; eHealth; mHealth

Background on Health Education Strategies

Health behavior change, which can be motivated by effective health education, has been called “our greatest hope for reducing the burden of preventable disease and death around the world” [1]. Improving maternal and child health through maternal education is especially important, given that more than 5 million children globally still die before reaching their 5th birthday, mostly from preventable illness [2]. Health education efforts aimed at promoting behaviors like breastfeeding and kangaroo mother care lie at the heart of many successful efforts to reduce

child mortality but increasing accessibility and dissemination of such health education remains challenging [3-5].

Prior Successful Strategies in Health Education

Narrative approaches to health education have emerged over the past decade as potentially powerful tools for promoting positive health behavior change [6,7]. As one of the longest-standing modes of human communication and knowledge transfer, Schank and Berman [8] suggest that stories function as cornerstones for learning, providing both the context for understanding and internalizing new information as well as a framework for remembering what we learn. Various theoretical models have been used to characterize the mechanisms by which narrative communication may lead to health behavior change, including the extended Elaboration Likelihood Model (ELM) [9]. In this model, the effectiveness of peripheral processing of the health messaging depends on the viewer's identification with the characters and their engagement (ie, absorption and transportation) with the storyline [9].

One particularly successful application of the narrative approach to health education can be seen in *entertainment-education* (E-E) [10]. E-E involves the delivery of health messages by embedding them in entertainment media. A substantial body of research suggests that E-E is an effective way of influencing beliefs, attitudes, and behaviors [9-13]. This approach may be even more effective in populations with lower motivation and ability to cognitively evaluate the health messages being delivered [10]. E-E media is particularly impactful when the following characteristics are incorporated: (1) an appealing storyline, (2) high-quality production, (3) unobtrusive persuasive messages, and (4) high potential for involvement with the characters.

Even the use of *simple pictures*, without an accompanying narrative storyline, has long been recognized as an effective way of enhancing attention to, and recall of, health education [14]. A significant body of prior research suggests that appropriate implementation of pictures can positively influence patient responses to health instructions and improve their health behaviors [14-16]. *Emerging technology* including mobile and online education platforms [3,5,17] present new opportunities for the dissemination of *video-based health education*. In recent years, researchers have begun to document the growing potential for online learning platforms, mobile phones, and tablets to fulfill important functions in the dissemination of open-access health education [17-20]. The increasing availability of mobile technology in low- and middle-income countries (LMIC) foretells promising new avenues for the delivery and scaling of mobile health education to communities who need it urgently [3,5,17,21,22].

Human-Centered Design of Global Health Education

Since 2013, a team of health educators at the Stanford University School of Medicine and the Stanford Center for Health Education has been grappling with a challenge: how do we create optimally effective and engaging video-based health education tools and deliver them at scale by adapting them across a broad spectrum of global learners? Over a 5-year period, our target audiences have ranged from medical students and practicing physicians to community health workers in LMIC and the general public. Dissemination pathways for our content include university learning management systems, open online learning platforms, mobile apps, government clinics, and nongovernmental health organizations.

The goal of creating content that is accessible, engaging, and relevant to the needs, wants, and experiences of the end user has been central to our learning trajectory. This has ultimately led us to apply a human-centered design (HCD) approach to the creation of our content and programs. HCD of health education revolves around several principles described in the literature, including empathy with the target communities, a process of rapid prototyping, feedback gathering and responsive iteration [23,24], as well as a tolerance for ambiguity and failure throughout the design process [25-27]. HCD involves integrating local characters, visuals, and narratives into accessible content, designed to deliver health messages that are collaboratively defined and created. A key principle of HCD is the idea that successful solutions should be created with the needs and wants of the end user in mind. Aligning with this principle, our content creation process has developed over time to rely increasingly on multiple cycles of feedback from community members in our target audience (our end users) and rapid iteration in response to that feedback. Especially in under-resourced and low-literacy settings, we have found that taking a community-based approach, by involving the target audience in the content creation process, is more likely to yield health education programs that are truly rooted in the contexts of the communities they intend to serve. Furthermore, through the creation of video-based teaching tools featuring narratives and visual elements that resonate with end users, we believe we can optimally engage our target audiences, empowering them to improve their own health behaviors and ultimately, their own health outcomes. The positive global response to our work contributed to the founding of the Stanford Center for Health Education in 2017 and its nonprofit global health education arm, the Digital Medical Education International Collaborative [28]. Overall, 8 video-based health and medical education initiatives, developed by our team over a 5-year period, generated the guidelines presented in this paper. These formative projects are described in Table 1 and presented along a development timeline map [29]. Multimedia Appendices 1-11 are referenced within this table and contain relevant video samples.

Table 1. Health and medical education initiatives (2013-2018).

Content subject matter	Technology, skill requirements, and scope	Audiences with delivery platforms, reach, and sample content	Collaborators, design approach, and feedback summary
Basic physiology and child health (2013)	 Basic digestion: sample physiology teaching video created in 2013 (Multimedia Appendix 1). Overall, 14 videos, 9 to 12 min each; amateur digital illustrations on black background; captured using screen-capture software (Camtasia); simultaneous audio capture via Rode microphone	Undergraduate students in a flipped-classroom Stanford Human Biology 121 course (content accessed by approximately 400 Stanford students via unlisted YouTube playlist [30]); learners in the general public (via The Khan Academy website [31] and The Khan Academy Medicine YouTube channel [32]), number of content views range from 43,524 (asthma [33])to 893,899 (basic metabolism [32])	Exploratory content; created by a single faculty member; modeled on Khan Academy teaching style; employed a purely didactic teaching approach; faculty content creator did not use a human-centered design approach; content was relatively cost-effective to produce: no professional illustrator was needed because basic editing was done by faculty content creator; students responded favorably to gradually unfolding visual elements, liked having the ability to pause, rewind, or watch at 2x speed if needed; black background wasted too much ink for students who wished to print out end screen as a study tool (feedback: public comments can be viewed here [32])
Infectious diseases (microbiology and immunology, 2014)	 Malaria teaching video created in 2014 for Stanford Medicine's infectious disease course (Multimedia Appendix 2). Overall, 36 videos, 8 to 10 min each; amateur digital illustrations on white background to facilitate printing of end screens; captured using Camtasia; asynchronous, scripted audio capture (via Rode microphone); subsequent synchronizing of audio and visual components in postproduction	International medical students as part of required microbiology/immunology course offerings (via university learning management systems and Stanford Medicine YouTube Playlist [34], 3000+ playlist views); international health sciences students (via Stories of Infection [35] open online course on Coursera, current enrollment: 17,116)	Content created in collaboration with faculty experts and medical students at Stanford University, University of California San Francisco; Duke University, University of Washington; and University of Michigan; formative feedback from medical students collected via focus groups throughout the course development period; medical student advisors attended weekly meetings, participated in script-writing and storyboarding process; medical student content users reported that the patient-centered narrative approach, interwoven with didactic elements, provided a useful framework for remembering the material and preferred a white background for easier printing of end screen as a study tool; (feedback from international Coursera learners: 99% positive, based on 4904 ratings)
Child nutrition and cooking (2014)	 Stanford's Child Nutrition and Cooking course was created in 2014 (Multimedia Appendix 3). Overall, 46 videos, 4 to 9 min each; amateur digital illustrations used in didactic videos + multimedia entertainment-education (E-E) cooking demonstrations; on-camera segments shot and edited by professional videographer	Undergraduate students in a flipped-classroom Stanford Human Biology 81 Q (content accessed via YouTube playlist [36]); international parents and caregivers (via the Stanford Child Nutrition and Cooking [37] open online course on Coursera), current enrollment: 315,834.	Stanford School of Medicine Faculty Nutrition Experts collaborated with parents and teachers at Stanford's Bing Nursery School; children and local celebrity chefs also featured to stimulate and empower parents to apply principles presented in nutrition education videos; parents responded favorably to positive role modeling, inclusion of children and celebrity chefs, and the connection made between theory and practice of child nutrition; adults without children requested a similar course on adult nutrition topics (feedback from international Coursera learners: 99% positive, based on 12,918 ratings)

Content subject matter	Technology, skill requirements, and scope	Audiences with delivery platforms, reach, and sample content	Collaborators, design approach, and feedback summary
Community maternal and child health (2015-18)	14 videos, 3 to 4 min each; digital illustrations using reveal animation (illustrations masked, then revealed to mimic live drawing, using Premier Pro software); illustration and editing support required; content adapted for: (1) South Africa in isiXhosa 2015  Xhosa nutrition in pregnancy video created for South African maternal child health promotion in 2015 (Multimedia Appendix 4) (2) India in Hindi 2017  Hindi nutrition in pregnancy video adapted for Indian maternal child health promotion in 2017 (Multimedia Appendix 5) (3) Burkina Faso in Dioula 2018 	Community health workers (CHWs) and their clients in (1) South Africa 2015: content [38] delivered via CHWs at the Philani Maternal Child Health and Nutrition Trust (Philani); India 2017: content [39] delivered via tablets and projector at Antara Foundation and Vivikenanda Tribal Hospital; Burkina Faso (content [40] pilot, 2018); app for Android devices created in 2015 [41]	Stanford School of Medicine Health Educators collaborated with CHWs and supervisors at the Philani Maternal Child Health and Nutrition Trust in Khayelitsha, South Africa; didactic and story-based maternal-child health education videos; illustrated by local artists and students; professional voiceover service used for the early videos: relatively expensive, language delivery did not consistently resonate with end users; 2016: began using the voices of CHWs and supervisors; content area priorities and scripts collaboratively developed and translated by Philani staff; Philani CHW end users responded favorably in a 2016 qualitative feasibility study [42]; feedback pending from adaptations created for India and Burkina Faso
Food and health for adults (2016)	 Stanford's Introduction to Food and Health course, created in 2016 (Multimedia Appendix 6). Overall, 28 videos, 4 to 6 min each; multimedia, E-E approach; video footage and animations; 10 on-camera cooking demonstrations (massive open online version only); filmmaker and professional animator engaged to support course creation	Stanford medical students as part of their nutrition unit (course content accessed via Stanford's university learning management system); practicing physicians via continuing medical education (CME) course (course content [43] accessed via Stanford Online CME platform); international learners (via YouTube playlist [44] and the Stanford Food and Health [45] open online course on Coursera, current enrollment: 183,229)	Stanford School of Medicine Faculty Nutrition Experts collaborated with a celebrity food journalist, local animators, and a food/lifestyle channel; video series integrated animation and multimedia, E-E; Course piloted by Stanford medical students in required nutrition block; students and international learners responded favorably to integration of celebrity food journalist's perspective and E-E approach; (feedback from international Coursera learners: 98% positive, based on 21,087 ratings)
Gender identity and children's health (2017)	 Health Across the Gender Spectrum (massive open online course) 2017 (Multimedia Appendix 7). Overall, 18 videos, 2 to 6 min each; amateur digitally illustrated videos; narrative approach interspersed with on-camera interviews; videographer and editor needed for live interviews	International learners (via Health Across the Gender Spectrum [46], open online course on Coursera, current enrollment: 12,668) and via YouTube playlist [47]; physicians via Stanford online CME course (pending release of in Oct 2018)	Stanford faculty at the School of Medicine collaborated with local stakeholders: Gender Spectrum, Planned Parenthood, the Stanford University Gender Clinic, Vaden Student Health Center; authentic narratives of transgender children, their parents, physicians, teachers, and 2 acclaimed transgender Stanford academics used to enhance engagement in the course content; interviewees (featured in the videos as animated line drawings, to protect their identities) played a formative role, through Rapid Iterative Testing and Evaluation [23,24], by sharing their stories and participating in the design of their illustrated characters; international learners expressed strong emotional responses to the personal narratives; (feedback from international Coursera learners: 98% positive, based on 1,680 ratings)

Content subject matter	Technology, skill requirements, and scope	Audiences with delivery platforms, reach, and sample content	Collaborators, design approach, and feedback summary
Breastfeeding promotion (2018)	100% Breastfed series launched in South Africa in 2018 (Multimedia Appendix 8). Overall, 4 parallel versions x10 videos, 2 to 4 min each; multimedia (live footage and illustrations); interviews with celebrity and community mothers; videographer, editor, professional illustrator, and logo designer involved	The general public (via Stanford's Short Course on Breastfeeding [48], an open online course on Coursera, current enrollment: 12,187) and via YouTube playlist [49]; South African mothers, caregivers, and health workers in multiple languages (via open-access, dedicated website [50] for this initiative)	Stanford Center for Health Education established its international education outreach arm, Digital Medical Education International Collaborative (Digital MEdIC) South Africa, with a local team based in Cape Town, led by the first author; video series created in collaboration with United Nations International Children's Emergency Fund; the Western Cape Department of Health; the National Department of Health in South Africa; as well as the University of Cape Town, Stellenbosch University and the Philani Maternal Child Health and Nutrition Trust; series features celebrities and community mothers, who participated in content creation by reacting to early drafts via WhatsApp; Philani CHWs responded positively to the inclusion of celebrities, multimedia approach, and branding of the course as well as its delivery using tablet devices, explored during "bodystorming" [23,51] (physical brainstorming sessions) described below. (Feedback from international Coursera learners: 99% positive, based on 1035 ratings)
National and regional government health education programs (completion in 2019)	Road to Health series content style: Narrative, icon style, maternal-child health library (75 videos planned, 2 to 3 min each); currently in production (6 videos completed); illustrator-animator and editor involved	The general public with a focus on mothers and caregivers, (dissemination via National Department of Health programs); see samples below: (1) Road to Health Trailer [52]; created for the South African Dept. of Health in 2018 (Multimedia Appendix 9); (2) Breastfeeding video [53]; created for the South African National Dept. of Health in 2018 (Multimedia Appendix 10); (3) Kangaroo Mother Care video [54] created for the South African National Dept. of Health in 2018 (Multimedia Appendix 11)	Digital MEdIC South Africa invited to collaborate with the Western Cape Department of Health, the National Department of Health, the DG Murray Trust, and Ilifa Labantwana to create content in support of South Africa's Road to Health Book [55]; initiative reaches every new mother who delivers in a health care facility; icon characters designed through a formative process involving Rapid Iterative Testing and Evaluation [23,24], to resonate across ethnic and socioeconomic demographics; story-based, community-narrated approach to the accompanying audio designed to maximize authenticity and identification with the characters; content will be disseminated through the National Department of Health programs

Adaptability Across Learner Populations, Platforms, and Time

An additional challenge when creating health education content for dissemination at scale is the need to easily adapt that content across different learner populations, different platforms, and across time. Adaptability across different learner populations, including language groups as well as cultural and socioeconomic contexts, is essential for broad scaling of health education content, both within countries and internationally. In community health education, this can mean designing characters that would resonate across ethnic and socioeconomic groups within a given country. In medical education, this can mean including both units of measurement or presenting a symptom generically as "an elevated body temperature" rather than "a fever of 102 degrees Fahrenheit," which limits the scalability of the content internationally. Easy adaptation facilitates the reach and therefore the impact of each video asset created. This also greatly contributes to the cost-effectiveness of such video-based health education.

The need to easily adapt content for existing and emerging platforms and delivery pathways also requires careful

consideration. Especially given the current global trend toward increasing penetration of smartphones and other mobile technology [56], video-based health education should ideally be optimized for mobile dissemination in the future, even if it is delivered in the shorter term using older platforms, such as university learning management systems, DVDs, offline tablets, or in-clinic television screens. As data costs decrease and the number of smartphones increase around the world, effective content needs to remain flexible and easily adaptable for emerging mobile dissemination pathways. This can include zooming in on diagrams or detailed images, such that the main teaching points can easily be appreciated even on smaller screens. Decisions about the nature of the visual media included in health education videos will also have enormous implications for file size, with the inclusion of video footage and high-resolution artwork resulting in much larger file sizes than more simple visual styles. Given the relatively high cost of data in some LMIC, file size can have significant implications in terms of adaptability for mobile push messaging.

Finally, because of the significant investment of cost and time in producing high-quality video content, this medium is best for disseminating health and medical educational content that will likely remain "evergreen" for several years. This implies selection of evidence-based concepts that are both foundational and unlikely to change considerably in the short term. For example, a health education video about the benefits of breastfeeding would likely have greater longevity than a video summarizing the differences in breastfeeding rates by country, as these are much more likely to change from year to year. Even subtle language choices can facilitate the longevity of video-based educational content. For example, "Until 2016, the World Health Organization categorically advised HIV positive mothers against mixed-feeding, regardless of their ARV [antiretroviral] treatment regimens" facilitates longevity better than "Until 2 years ago, the World Health Organization categorically advised HIV positive mothers against mixed-feeding, regardless of their ARV [antiretroviral] treatment regimens."

South Africa as a Case Study

In South Africa (SA), a country with 11 national languages and a population that is ethnically, culturally, and socioeconomically diverse, a community-based, HCD approach has been instrumental in addressing the challenges of engagement and adaptability. The 2 projects described here illustrate the application of such an approach.

The 100% Breastfed Initiative

The 100% Breastfed Initiative, launched in March 2018, consisted of a series of 10 breastfeeding educational videos targeted primarily toward SA mothers. During the formative, early discussions with local health workers and community members, it was suggested that the educational series should include on-camera interviews with celebrity mothers who would serve as role models and representatives of SA's ethnic diversity. Feedback from 2014 and 2016 massive open online courses (Table 1) suggested that the inclusion of the celebrity perspective seemed to enhance engagement within the general public. Overall, 3 SA celebrity mothers were identified, interviewed, and included in early draft videos.

Community members and local clinics viewed the content and suggested adding the perspectives of community mothers to emphasize that many of the challenges faced by new mothers in SA transcend these demographics. This suggestion was accepted and implemented in the next iteration of the video series. Freehand, digital illustrations, drawn by a local SA artist, and branding (including a logo, see Table 1) were later also integrated in response to input from a large SA women's media publishing house. The illustrations were intended to underscore the cross-cultural nature of the health messages, by reflecting South Africa's ethnic diversity. The logo and branding were intended to facilitate national dissemination of the campaign.

A total of 4 versions of the course were created in parallel for different audiences between July 2017 and March 2018 (English for SA mothers, isiXhosa for SA mothers, English for SA health workers, and an international version, launched on Coursera). The scripts for each teaching video were created in real-time collaboration with 11 local stakeholders using Google Drive to manage version control. For each video, an audio-visual script [57] showing 3 parallel versions would be finalized before production would begin. Version_2 (advanced) was later converted into 2 versions, one narrated by twin SA medical doctors, Vela and Phinda Njisane, for SA health workers and the fourth version narrated by a Canadian health educator, for a massive open online audience.

As early drafts of the videos were completed, these were shared with local community advisors using WhatsApp, a communication app that supports the transfer of video files. This feedback led to several iterations of the images, characters, and content, a process that continued throughout the production period. WhatsApp was a significant production asset, useful in soliciting feedback from community members who did not regularly use email and did not feel comfortable editing scripts using Google Drive.

Final content prototypes were loaded onto android tablets to be used by Community Health Workers in physical brainstorming sessions, described as "bodystorming" [23,51] in prior HCD literature. During these sessions, community health workers engaged in role-play that involved testing the intervention in mock-counseling sessions, observed by members of the production team.

The final video series was launched in March 2018 and has since also been translated and dubbed in Afrikaans. Over the next 12 months, the content will be translated into several other national languages.

A cluster randomized controlled trial involving 1008 pregnant women was launched in September 2018 to study the effects of the 100% Breastfed series on breastfeeding rates in the study participants.

Although the videos have been well received to date (see Table 1), we realized that this approach had some limitations. First, the file size of videos containing live footage is considerably larger than those containing simple illustrations or animations. This makes dissemination via mobile push messaging to

smartphones prohibitive at present, due to the relatively high cost of data in South Africa. Second, adapting the content for use in other African countries would require identifying and interviewing local mothers in each country, then re-editing the series to give it a local flavor. Although the freehand, digital illustrations are somewhat easier to adapt, they would also require redrawing for each target country to accurately represent different styles of dress, local facial features, local foods, etc. We encountered similar issues when adapting the 2015 Community Maternal and Child Health course (originally created for an under-resourced urban South African population) for audiences in India (2017) and Burkina Faso (2018). Although live footage and realistic depictions of characters can be powerful ways to enhance visual identification with characters, this approach also limits the ease of adaptation of the resulting content, an especially important consideration in countries or geographic regions that are home to multiple diverse ethnic and socioeconomic groups.

The Road to Health Initiative

The Road to Health Book (RTHB) is a child health initiative of the South African National Department of Health, first launched in 1973 and updated for relaunch in 2018. The goal of the initiative is to provide new mothers across South Africa with education and resources to help them become more engaged in their child's health, in partnership with their child's health care providers.

In 2018, we were invited to begin creating video-based content to support the RTHB initiative. The main challenges were (1) to create content that would resonate across SA and (2) to dramatically reduce the file size of the video content because the data charges associated with using the content would be reverse billed to the National Department of Health.

Formative discussions with local SA stakeholders in maternal-child health, as well as feedback from earlier online courses, helped to distill a central goal: to create a visual style that was extremely simple (yielding smaller video file sizes) and universal enough to resonate across ethnic and socioeconomic groups, ideally with the possibility of repurposing the content for a global audience in a cost- and time-effective way.

While retaining visual simplicity, the content needed to elicit enough emotion in the viewer to engage a lay audience who was choosing to watch the content voluntarily (ie, this was not required viewing as part of a formal educational program). Feedback from a 2017 massive open online course (Table 1) suggested that the inclusion of dialogues between family members or individuals involved in close interpersonal relationships was a powerful way to engage learners, even in topic areas where initial motivation to attend to the content was low.

Over the following 4 months, we solicited feedback on different iterations of simple icon-style characters through a process of rapid iterative testing and evaluation described in the HCD literature [23,24]. Each resulting prototype served as the new design direction until consensus was reached on a modified version of the "universal human" icon that is used globally in public places, most commonly to indicate the location of restroom facilities. Table 2 summarizes the prototyping pathway. Visually, these icon-style characters were enhanced through the use of various nonskin tone colors and textures. Their settings were kept similarly universal by relying on simple shapes and lines to represent props in the environment, such as eating surfaces or places to sit.

Table 2. Visual development progression.

Visual style development	Visual considerations and feedback
	Live footage of interviews with community members and local celebrity mothers used to enhance engagement in the 100% Breastfed Health Education series; changing voiceovers allowed for in-country adaptation for different learner populations but adapting the course for other African countries would require significant resources as on-camera interviews would need to be rerecorded for each new country
	Freehand digital illustrations by local illustrator featured South African (SA) mothers and families; characters designed to represent varying demographics to enhance identification with the content broadly across SA; each teaching video incorporated different characters, underscoring the common experiences and challenges faced by new mothers, related to infant feeding
	Icon-style concept art (iteration 1) developed to elicit feedback on the icon approach; visual style not favorably received by community members and local stakeholders: (1) Characters appeared to belong to 1 ethnic group due to skin color and hair (with limited adaptability for non-Western audiences) and (2) clothing styles were also thought to be unacceptable for many non-Western audiences
	New icon-style concept art (iteration 2) developed; visual style favorably received by community members and local stakeholders, with some limitations: (1) textures and simplicity of images found to be acceptable, with neutral background and concept of *universal surfaces* found to be preferable to specific props like table/chairs and (2) clothing style and visible ethnic cues like skin color and hair could limit ease of adaptability for other cultures
	New icon-style concept art (iteration 3) developed; visual style somewhat favorably received by community members and local stakeholders, with some limitations: (1) simple icon characters thought to be more easily generalizable across different cultures and demographics and (2) early feedback suggested that images may be too juvenile (“cartoony”) for adult learners and might too closely resemble emojis
	New icon-style concept art (iteration 4) developed; prototype resonated with national government and nongovernmental organizations as well as community advisors: (1) neutral background incorporating *universal surfaces* from iteration 2 thought to be appropriate for SA due to subtle Ndebele (African painting style) through a double black line; (2) lack of visible ethnic or socioeconomic identifiers found to be an asset; (3) additional benefit: colors could be adapted as needed to match various health campaigns
	New icon-style concept art (iteration 5) developed; prototype approved for use in the Road to Health Book initiative by the National Department of Health and resonated with community advisors: (1) prototype incorporates *universal surfaces* neutral background and character design free of ethnic and socioeconomic identifiers, (2) color scheme easily altered as needed, (3) drop-shadowing makes texture icons appear somewhat three-dimensional (without adding to production costs) when paired with color, animation, and voiceovers

Localization of this new video content style relies almost entirely on the audio components, which feature the voices and narratives of the target audience as well as a soundtrack. By recruiting the end users of the content to help shape and deliver the voiceovers, these videos acquired an authentic, community-based quality, according to early, informal feedback from SA community members.

The next steps will include conducting formal focus groups across different language groups as part of the National Department of Health’s evaluation of the RTHB initiative.

The main advantage, from a production perspective, of employing a simple visual style with reliance on the voiceover and narrative for localization, is that this approach is easily adaptable, by changing the audio components only. Furthermore, involving the end user (or target community) in the voiceover recordings and intervention development can serve to empower these communities and facilitate ownership of both the educational content and the health behaviors they aim to promote.

Theoretical Framework

This work rests upon the theoretical underpinnings of the ELM [58], which suggests that there are 2 contributing pathways to achieve the changes in attitude that predict a desired behavioral outcome. The first *central route* is influenced by the learners’ motivation and ability to cognitively process the information presented. This can be influenced by factors such as the length of the content and the degree to which the language used is accessible to the learner. The second *peripheral route* relies on cues embedded in the method of information delivery that contribute to its relative acceptability to the learner. Positive, peripheral cues, such as the learners’ subjective evaluation of the narrator or the learners’ emotional involvement in the content, lead to peripheral attitude changes, which, although less enduring than central attitude changes, can positively influence the learners’ motivation to process the messaging via the central route [58].

Production Guidelines

In this section, we outline the steps we have followed when creating video-based health education content, using HCD. These production guidelines help to standardize the video production process across projects and teams as well as supporting quality control and time management. The guidelines were developed collaboratively, over 12 months, by content creation leads and creative team members in the United States and South Africa. The guidelines are based on prior research on the effects of different video production strategies on learner engagement [59]. To some extent, the desired style of content

being produced will define the video production workflow [60], as well as the production team skill sets and the budget needed to support such work. For example, creation of the storyboarding and animatic steps, described below, would primarily be text-based descriptions of anticipated footage for a video integrating mostly camera footage, whereas the storyboard and animatics would primarily consist of draft illustrations for an animated video. In Table 3, we describe our workflow as well as provide examples and strategies for optimizing each phase. Multimedia Appendices 11-14 are referenced within this table and contain relevant video samples.

Takeaways

Applying a community-based, HCD approach during the collaborative development of scalable video-based health education content can yield significant benefits including:

- Target communities are empowered as the strengths and resources within them are identified, harnessed, and showcased. Recognizing the human resources already available within a community, by involving them in the production of their own health education content, serves to validate, educate, and enable that community. Empowered community members can promote health within their peer groups and can advocate for increased access to resources like health and social services on behalf of their communities. An empowered community recognizes and prioritizes health education and the behaviors associated with improved outcomes. They also become valuable partners in the dissemination of their own health education content.
- This approach fosters improved health awareness, increases identification of the target communities with the content, and creates a sense of ownership among community members who participated in the production of that content. General health awareness improves as a result of peer-to-peer education (word-of-mouth) between those involved in the content creation and their fellow community members. A sense of ownership results from the fact that time and energy have been invested by the community in the content creation process. Identification is enhanced by the fact that content better reflects the context, needs, and wants of the target community. An increased sense of awareness, ownership, and identification with the content decreases resistance to health messages (also called counterarguing) as the messages are perceived as coming from sources that have been internally validated rather than being foreign and external to the community.
- Localization through the incorporation of local narratives and simple visuals yields content that can be adapted for other audiences in a cost- and time-efficient manner. Incorporating local stories and the voices of community members results in content that aligns desired health messages with local beliefs, perspectives, and language used to communicate around these issues. As such, content relies more on audio than visual components for its cultural resonance and can be easily adapted for other audiences—often simply by changing the narrative voiceover. Easier adaptation, in turn, facilitates scaling of health education to other regions. Furthermore, simplifying visual components reduces the financial burden associated with transferring large data files.

Table 3. Production workflow.

Main objectives	Description	Example
Instructional design phase	Define target audience for final video (1) identify general intent/purpose of the video for each audience, (2) specify learning objectives, (3) decide on the target length of video, (4) define visual style of video, (5) define pedagogic approach (ie, narrative, didactic, or hybrid), (6) create a design document [61]	The Grow Great design document [62] was created in preparation for a collaboratively developed growth stunting reduction campaign. Video content developed in collaboration with the DG Murray Trust, South Africa.  Sample template header for design document
Drafting of an audio-visual script	Consider how the learner will be guided through the learning objectives, including which background information may need to be provided; draft core dialogue if the video will be "scripted" or record interviews if teaching scripts will be derived from recorded interviews.	Audio-visual script [63] for Grow Great Video 1. Prescripted teaching video [54]; (Multimedia Appendix 11); teaching video derived from a recorded interview [64] (Multimedia Appendix 12)  Audio-visual script with storyboarding
Storyboarding	Plan accompanying visuals for each frame (each sentence or main point of the video); generate rough sketches of illustrations to be used; obtain video footage or other images if relevant	Audio-visual script [63] (with storyboarding) for Grow Great Video 1
Obtaining feedback from stakeholders and end users	Share A/V script with relevant local stakeholders and selected creative advisors within the target community. (Feedback can be collected via Google Drive, videoconference with screen sharing or Word documents by email with "track changes" to manage version control); mobile communication tools that support video transfers (such as WhatsApp) are valuable for soliciting feedback from target community members who do not regularly use email.	 Feedback session in USA and South Africa via Zoom videoconference with screen sharing
Audio recording	Identify community members from within the target community to record the voiceover for the video. (This process may involve *auditioning* interested community members, with external feedback from other end users informing the selection of finalists); consider that the quality of the audio recording can vary depending on the available recording equipment and location. If resources are limited, a voice-recording app on a smartphone can be used to record audio scripts inside a closed car, parked in a quiet area. Smaller, insulated spaces can be adapted to yield clean audio with little or no echo (ie, closets containing clothing or blankets hung on walls to absorb sound waves). If available, a soundproofed recording booth with a high- quality microphone is preferable; community members may need *coaching* to deliver lines in a way that sounds authentic and compelling. (Tips: encouraging community voice performers to stand while recording and use facial expressions, such as smiling and raising eyebrows, can yield a more convincing delivery.) Local celebrities can be recruited as compelling voice talent for public health education interventions.	 Makeshift "Recording Studio" In Khayelitsha, South Africa  Virtual recording session with community voice talent, Qhawe Nkopane (age 10 years)  South African celebrity Zolani Mahola (lead singer of the band *Freshlyground*) volunteered to narrate the 100% Breastfed series, Cape Town, 2018
Animatics (creating sequences of images or sketches)	Combine still frames of storyboard with recorded audio using editing software such as Camtasia, Premiere Pro, or others; use video collaboration review tools such as Frame.io for internal review and feedback from local stakeholders; use WhatsApp or face-to-face meetings to gather feedback from end users who are more comfortable with these feedback avenues; use *bodystorming* sessions to role-play the use of the intervention in the field by end users.	 Frame.io [65] is an online video review and collaboration tool. An "animatic" [66] (Multimedia Appendix 13) is a draft video created using rough sketches only.  Community health workers engage in role-play exercises to test the use of the intervention for future home visits with new mothers. (Simple props such as the "baby" in this scenario can help with early identification of potential barriers to implementation.)

Main objectives	Description	Example
Art assets	Digital illustration tools such as Photoshop or Adobe Illustrator are useful; if animations will be incorporated, specify the naming conventions for layers and individual elements to be animated, for example: "character_bodypart" (eg, girl_head, girl_body, girl_armL, girl_armR, girl_legL)	Digital illustration tools such as Photoshop or Adobe Illustrator can be used to create simple images that underscore teaching messages and enhance engagement.
Animation (art assets edited to create the illusion of movement)	If animation used, *After Effects* and/or *Premiere Pro* can be useful software assets; use animatic as a guide layer.	An "animatic" (Multimedia Appendix 13) can be used as a guide layer for animation.
Delivery	Render (export) videos for high definition and mobile viewing; upload to Frame.io and Google Drive for sharing and a final round of feedback.	Growth stunting awareness video created with the DG Murray Trust, South Africa, 2018 [67] (Multimedia Appendix 14)
Archiving	Name final videos according to consistently defined conventions (ie, FINAL_ Grow Great_1_July18); file video creation documents and all iterations for future look-back; file art, music, and sound assets in a production library for reuse in future projects.	

Human beings are hardwired to learn through stories and personal narratives. By collaborating with target communities to (1) identify and shape relevant narratives, (2) harness local talent and expertise to convey priority health messages through them, and (3) "digitally package" the resulting content for easy, open-access global scaling, we can marry the oldest and newest forms of teaching in a synergistic and innovative approach to health education.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Basic digestion: sample physiology teaching video created in 2013.

[MP4 File (MP4 Video), 19MB - jmir_v21i1e12128_app1.mp4]

Multimedia Appendix 2

Malaria teaching video created in 2014 for Stanford Medicine's infectious disease course.

[MP4 File (MP4 Video), 14MB - jmir_v21i1e12128_app2.mp4]

Multimedia Appendix 3

Stanford's child nutrition and cooking trailer was created in 2014.

[MP4 File (MP4 Video), 27MB - jmir_v21i1e12128_app3.mp4]

Multimedia Appendix 4

Xhosa nutrition in pregnancy video created for South African maternal child health promotion in 2015.

[MP4 File (MP4 Video), 10MB - jmir_v21i1e12128_app4.mp4]

Multimedia Appendix 5

Hindi nutrition in pregnancy video adapted for Indian maternal child health promotion in 2017.

[MP4 File (MP4 Video), 5MB - jmir_v21i1e12128_app5.mp4]

Multimedia Appendix 6

Stanford's Introduction to Food and Health trailer, created in 2016.

[MP4 File (MP4 Video), 20MB - jmir_v21i1e12128_app6.mp4]

Multimedia Appendix 7

Trailer for health across the gender spectrum (massive open online course) 2017.

[MOV File, 15MB - jmir_v21i1e12128_app7.mov]

Multimedia Appendix 8

100% Breastfed trailer, launched in South Africa in 2018.

[MP4 File (MP4 Video), 9MB - jmir_v21i1e12128_app8.mp4]

Multimedia Appendix 9

Road to health trailer created for the South African Department of Health in 2018.

[MP4 File (MP4 Video), 7MB - jmir_v21i1e12128_app9.mp4]

Multimedia Appendix 10

Breastfeeding video created for the South African National Department of Health in 2018.

[MP4 File (MP4 Video), 30MB - jmir_v21i1e12128_app10.mp4]

Multimedia Appendix 11

Kangaroo mother care video created for the South African National Department of Health in 2018.

[MP4 File (MP4 Video), 7MB - jmir_v21i1e12128_app11.mp4]

Multimedia Appendix 12

Teaching video (derived from a recorded interview) created for open online course: Health Across the Gender Spectrum.

[MP4 File (MP4 Video), 11MB - jmir_v21i1e12128_app12.mp4]

Multimedia Appendix 13

Sample animatic.

[MP4 File (MP4 Video), 6MB - jmir_v21i1e12128_app13.mp4]

Multimedia Appendix 14

Growth stunting awareness video created with the DG Murray Trust, South Africa, 2018.

[MP4 File (MP4 Video), 17MB - jmir_v21i1e12128_app14.mp4]

References

1. Glanz K, Rimer BK, Viswanath K. Health behavior and health education: theory, research, and practice. Hoboken, NJ: John Wiley & Sons; 2008.
2. World Health Organization. World health statistics 2016: monitoring health for the SDGs sustainable development goals. In: World health statistics 2016. Geneva: World Health Organization; 2016.
3. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14:188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
4. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
5. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17(Suppl 1):82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]
6. Hinyard LJ, Kreuter MW. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. Health Educ Behav 2007 Oct;34(5):777-792. [doi: 10.1177/1090198106291963] [Medline: 17200094]

7. Shaffer VA, Focella ES, Hathaway A, Scherer LD, Zikmund-Fisher BJ. On the usefulness of narratives: an interdisciplinary review and theoretical model. Ann Behav Med 2018 Apr 19;52(5):429-442. [doi: 10.1093/abm/kax008] [Medline: 29684135]
8. Schank RC, Berman TR. The pervasive role of stories in knowledge and action. In: Narrative impact: Social and cognitive foundations. Oxfordshire: Psychology Press; 2003:287-288.
9. Slater MD, Rouner D. Entertainment education and elaboration likelihood: understanding the processing of narrative persuasion. Commun Theory 2002;12(2):173-191. [doi: 10.1111/j.1468-2885.2002.tb00265.x]
10. Slater MD. Entertainment education and the persuasive impact of narratives. In: Green MC, Strange JJ, Brock TC, editors. Narrative impact: Social and cognitive foundations. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2002:157-181.
11. Rogers EM, Vaughan PW, Swalehe RM, Rao N, Svenkerud P, Sood S. Effects of an entertainment-education radio soap opera on family planning behavior in Tanzania. Stud Fam Plann 1999 Sep;30(3):193-211. [Medline: 10546311]
12. Shen F, Han J. Effectiveness of entertainment education in communicating health information: a systematic review. Asian J Commun 2014;24(6):616-617. [doi: 10.1080/01292986.2014.927895]
13. Wilson KE, Beck VH. Entertainment outreach for women's health at CDC. J Womens Health Gend Based Med 2002 Sep;11(7):575-578. [doi: 10.1089/152460902760360522] [Medline: 12396890]
14. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns 2006 May;61(2):173-190. [doi: 10.1016/j.pec.2005.05.004] [Medline: 16122896]
15. Delp C, Jones J. Communicating information to patients: the use of cartoon illustrations to improve comprehension of instructions. Acad Emerg Med 1996 Mar;3(3):264-270 [FREE Full text] [Medline: 8673784]
16. Katz MG, Kripalani S, Weiss BD. Use of pictorial aids in medication instructions: a review of the literature. Am J Health Syst Pharm 2006 Dec 01;63(23):2391-2397. [doi: 10.2146/ajhp060162] [Medline: 17106013]
17. 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] [Medline: 27144393]
18. World Health Organization. World Health Organization. Geneva: World Health Organization; 2011. New horizons for health through mobile technologies: second global survey on eHealth URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2018-11-28] [WebCite Cache ID 74GOnJrgk]
19. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
20. Adam M, Young-Wolff KC, Konar E, Winkleby M. Massive open online nutrition and cooking course for improved eating behaviors and meal composition. Int J Behav Nutr Phys Act 2015 Dec 03;12:143-144 [FREE Full text] [doi: 10.1186/s12966-015-0305-2] [Medline: 26630879]
21. Källander K, Tibenderana JK, Akpogheneta OJ, Strachan DL, Hill Z, ten Asbroek AH, et al. Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low- and middle-income countries: a review. J Med Internet Res 2013 Jan 25;15(1):e17 [FREE Full text] [doi: 10.2196/jmir.2130] [Medline: 23353680]
22. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med, 2013;10(2):e1001382-e1001383 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
23. Hanington B, Martin B. Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Beverly, MA: Rockport Publishers; 2012.
24. Medlock MC, Wixon D, Terrano M, Romero R, Fulton B. Semantic Scholar. 2002. Using the RITE method to improve products: a definition and a case study URL: https://pdfs.semanticscholar.org/5340/ef8a91900840263a4036b0433a389b7097b2.pdf?_ga=2.62812652.1990280664.1543398844-1967781214.1525692408 [accessed 2018-11-28] [WebCite Cache ID 74GOzwlbm]
25. Bazzano AN, Martin J, Hicks E, Faughnan M, Murphy L. Human-centred design in global health: a scoping review of applications and contexts. PLoS One 2017;12(11):e0186744 [FREE Full text] [doi: 10.1371/journal.pone.0186744] [Medline: 29091935]
26. Brown T, Wyatt J. Development Outreach. 2010. Design thinking for social innovation URL: http://siteresources.worldbank.org/WBI/Resources/213798-1278955272198/Design_Thinking_for_SocialInnovation2.pdf [accessed 2018-11-28] [WebCite Cache ID 74GP8xxBc]
27. Kolko J. Harvard Business Review. 2015. Design thinking comes of age URL: https://enterprisersproject.com/sites/default/files/design_thinking_comes_of_age.pdf [accessed 2018-11-28] [WebCite Cache ID 74GPCDdbi]
28. Prober C. The Digital Medical Education International Collaborative. 2017. URL: https://www.digitalmedic.org [accessed 2018-11-24] [WebCite Cache ID 725eY2TwH]
29. Kontinentalist Pte Ltd. Digital MEdIC Map by Kontinentalist. 2018. URL: https://storage.googleapis.com/storymap/dmstory.html [accessed 2018-11-21] [WebCite Cache ID 725g1QxXf]
30. Adam M. YouTube. 2013. Stanford HB121 Child Health Class Videos URL: https://www.youtube.com/playlist?list=PLdzGMzuJfqWkVDSDxe7wGj9LZMge3RWyn [accessed 2018-11-25] [WebCite Cache ID 725h8hdHv]

31. Adam M. Khan Academy. 2013. URL: https://www.khanacademy.org/partner-content/stanford-medicine/growth-and-metabolism/v/digesting-food?ref=stanford_staff_picks [accessed 2018-11-24] [WebCite Cache ID 727Lmkvn3]
32. Adam M. YouTube. 2013. Khan Academy Medicine: Basics of Metabolism, Stanford Medicine Partner Content URL: https://www.youtube.com/watch?v=wQ1QGZ6gJ8w [accessed 2018-11-24] [WebCite Cache ID 727MOK9se]
33. Adam M. YouTube. 2013. Khan Academy Medicine: Asthma, Stanford Medicine Partner Content URL: https://www.youtube.com/watch?v=SGKHQqHXmpQ [accessed 2018-11-23] [WebCite Cache ID 727MkdWgj]
34. Adam M. YouTube. 2014. Springboard Video Series, The Robert Wood Johnson Reimagining Medical Education Initiative URL: https://www.youtube.com/playlist?list=PLkdHelTlDZnETkBr6t7XVk0l2jzOxnBS1 [accessed 2018-11-23] [WebCite Cache ID 727MwvLAK]
35. Adam M. Coursera. 2017. Stories of Infection URL: https://www.coursera.org/learn/stories-of-infection [accessed 2018-11-23] [WebCite Cache ID 72DBgl3MX]
36. Adam M. YouTube. 2014. Stanford Child Nutrition and Cooking URL: https://www.youtube.com/watch?v=dcaXZ0fGB0U&list=PLB0AYRZmsMalvNbfs1qMWUeB2YF3CcOgI [accessed 2018-11-24] [WebCite Cache ID 727HXxY4Z]
37. Adam M. Coursera. 2014. Stanford Child Nutrition and Cooking Course URL: https://www.coursera.org/learn/childnutrition [accessed 2018-11-24] [WebCite Cache ID 72D9uE7y2]
38. Adam M. YouTube. 2015. Maternal Child Health: South Africa URL: https://www.youtube.com/playlist?list=PLB0AYRZmsMalxdR_wB1YCQ8o8A5vok6zf [accessed 2018-11-24] [WebCite Cache ID 72DAiBLW9]
39. Adam M. YouTube. 2017. Maternal Child Health: India URL: https://www.youtube.com/watch?v=1xEdb1cBd7Q&list=PLB0AYRZmsMaltYYzgryYpZr2Qfz9D9dNS [accessed 2018-11-24] [WebCite Cache ID 72DAyZeaV]
40. Adam M. YouTube. 2018. Maternal Child Health: Burkina Faso URL: https://www.youtube.com/watch?v=w2yNE6YKicw&list=PLB0AYRZmsMaksuy2MS3WkmgsNgZwNmWLy&index=6 [accessed 2018-11-24] [WebCite Cache ID 72DB5lvLG]
41. Adam M. Stanford Maternal Child Health. 2015. Stanford Health Outreach App for Android URL: https://play.google.com/store/apps/details?id=health.timtully.com.stanford_health_app [accessed 2018-11-25] [WebCite Cache ID 72DF5nC4H]
42. Coetzee B, Kohrman H, Tomlinson M, Mbewu N, Le Roux I, Adam M. Community health workers' experiences of using video teaching tools during home visits-A pilot study. Health Soc Care Community 2018 Mar;26(2):167-175. [doi: 10.1111/hsc.12488] [Medline: 28872210]
43. Adam M. Stanford Center for Continuing Medical Education. 2016. Stanford Introduction to Food and Health CME URL: https://www.youtube.com/playlist?list=PLuP5qOnCjJ-4Vg6vb-r7mi5HPMLiVJtpl [accessed 2018-11-24] [WebCite Cache ID 72DC3Tw0g]
44. Adam M. YouTube. 2016. Stanford Introduction to Food and Health URL: https://www.youtube.com/watch?v=z7x1aaZ03xU&list=PLB0AYRZmsMake5dK42lB2hkSlRgpDVsWl [accessed 2018-11-23] [WebCite Cache ID 72DC7uO5a]
45. Adam M. Coursera. 2016. Stanford Introduction to Food and Health URL: https://www.coursera.org/learn/food-and-health [accessed 2018-11-23] [WebCite Cache ID 72DC9bvsG]
46. Adam M. Coursera. 2017. Health Across the Gender Spectrum URL: https://www.coursera.org/learn/health-gender-spectrum [accessed 2018-11-28] [WebCite Cache ID 74GPXo4zK]
47. Adam M. YouTube. 2017. Health Across the Gender Spectrum URL: https://www.youtube.com/watch?v=FBdwQsTY7ec&list=PLB0AYRZmsManwPn7dI9voNuQCifj3khtY [accessed 2018-11-23] [WebCite Cache ID 72DUFfWOR]
48. Adam M. Coursera. 2018. Stanford's Short Course on Breastfeeding URL: https://www.coursera.org/learn/breastfeeding [accessed 2018-11-23] [WebCite Cache ID 72DUWvBsx]
49. Adam M. YouTube. 2018. 100% Breastfed URL: https://www.youtube.com/playlist?list=PLB0AYRZmsMamYsa29kdwQA9THyq4owm-c [accessed 2018-11-22] [WebCite Cache ID 72DV3RaBe]
50. Adam M. 100 Percent Breastfed. 2018. URL: http://www.100percentbreastfed.co.za [accessed 2018-11-23] [WebCite Cache ID 72DVDRHdH]
51. Burns C, Dishman E, Verplank W, Lassiter B. Actors, hairdos & videotape: Informance design. 1994 Presented at: CHI '94 Conference Companion on Human Factors in Computing Systems; April 24 - 28, 1994; Boston, Massachusetts, USA.
52. Adam M. YouTube. 2018. SA National Department of Health URL: https://youtu.be/__xRxgFOI98 [accessed 2018-11-24] [WebCite Cache ID 72DXIIG6R]
53. Adam M. YouTube. 2018. South African Road to Health Book: SA National Department of Health URL: https://www.youtube.com/watch?v=yAP_jFAY3Ms&feature=youtu.be [accessed 2018-11-25] [WebCite Cache ID 72DXV7zBE]
54. Adam M. Kangaroo Mother Care. 2018. South African Road to Health Book: SA National Department of Health URL: https://youtu.be/iqZkrAhm-Hk [accessed 2018-11-25] [WebCite Cache ID 72DmhDF30]
55. South African National Department of Health. Road to Health. URL: https://roadtohealth.co.za/about-us/ [accessed 2018-11-25] [WebCite Cache ID 72Dmvaz7q]

56. Kratzke C, Cox C. Smartphone technology and apps: rapidly changing health promotion. Glob J Health Educ Promot 2012;15(1):15-16.
57. Adam M. Google. 2017. Breastfeeding Sample Audio-Visual Script URL: https://docs.google.com/document/d/1IY3rJ1xqnDWRAx9MiY37ewgkccaCyR_6RajVrNa9gaU/edit [accessed 2018-11-25] [WebCite Cache ID 72Dn96j5N]
58. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. Adv Exp Soc Psychol 1986;19:123-205. [doi: 10.1016/S0065-2601(08)60214-2]
59. Guo PJ, Kim J, Rubin R. How video production affects student engagement: an empirical study of MOOC videos. 2014 Presented at: L@S '14 Proceedings of the first ACM conference on Learning; March 04 - 05, 2014; Atlanta, Georgia, USA p. 41-50.
60. Johnston S, Walker K. Digital MEdIC Video Production Workflow. 2018. URL: https://docs.google.com/document/d/1Tx9AARsmJDsV_3MPWLFKpnYmdl-ZDedM_pXVZ_nDfYY/edit#heading=h.moe3skrvlh63 [accessed 2018-11-25] [WebCite Cache ID 72DnMXx7B]
61. Johnston S, Walker K. Digital MEdIC Design Document Template. 2018. URL: https://docs.google.com/document/d/1ydmrFdwzrkTHfPOQGTLTpJeUdyfS4MiQLmBYAxOJxsY/edit [accessed 2018-11-25] [WebCite Cache ID 72DnmQtvV]
62. Johnston S, Walker K, Adam M. Grow Great Design Document. 2018. URL: https://docs.google.com/document/d/1GjJEe2DpIArCC-cHLswcH0z4Z-gC37ZFmUrxrx839Xk/edit [accessed 2018-11-25] [WebCite Cache ID 72DnbZ5Ct]
63. Johnston S, Walker K, Adam M. Grow Great Audio-Visual Script Video 1. 2018. URL: https://docs.google.com/document/d/1vd_hW1dSOgSTs-tr57ZU4oTxQ5gersipNA2UavGVDoQ/edit#heading=h.f0fz14s9j3j7 [accessed 2018-11-25] [WebCite Cache ID 72Dnxqppt]
64. Adam M. You Amaze me Every Day. 2017. Sample video: Health Across the Gender Spectrum URL: https://youtu.be/aDihlx776vM [accessed 2018-11-25] [WebCite Cache ID 72FTV1k7L]
65. Frame.io. 2016. URL: https://frame.io/ [accessed 2018-11-24] [WebCite Cache ID 72DoHHhWo]
66. Johnston S, Sanchez C. Animatic Grow Great Video 1. 2018. URL: https://www.youtube.com/watch?v=D_Zt561GP2A&feature=youtu.be [accessed 2018-11-25] [WebCite Cache ID 72DobzP0a]
67. Adam M, Eley N, Johnston S, Sanchez C, Kirton A. YouTube.: Stanford Center for Health Education and the DG Murray Trust; 2018. Grow Great Campaign URL: https://www.youtube.com/watch?v=Eddb4_rI1Fw&feature=youtu.be [accessed 2018-11-24] [WebCite Cache ID 72DoxsUPA]

Abbreviations

CHW: community health worker
CME: continuing medical education
E-E: entertainment-education
ELM: Elaboration Likelihood Model
HCD: human-centered design
LMIC: low- and middle-income countries
RTHB: Road to Health Book

Edited by G Eysenbach; submitted 07.09.18; peer-reviewed by K Bond; comments to author 14.10.18; accepted 19.10.18; published 30.01.19.

Please cite as:
Adam M, McMahon SA, Prober C, Bärnighausen T
Human-Centered Design of Video-Based Health Education: An Iterative, Collaborative, Community-Based Approach
J Med Internet Res 2019;21(1):e12128
URL: http://www.jmir.org/2019/1/e12128/
doi:10.2196/12128
PMID:30698531

Original Paper

Web-Based Exercise as an Effective Complementary Treatment for Patients With Nonalcoholic Fatty Liver Disease: Intervention Study

Daniel Pfirrmann[1], PhD; Yvonne Huber[2], MD; Jörn Markus Schattenberg[2], MD; Perikles Simon[1], MD, PhD

[1]Department of Sports Medicine, Disease Prevention and Rehabilitation, Institute of Sports Science, Johannes Gutenberg University, Mainz, Germany
[2]I. Department of Medicine, University Medical Center, Johannes Gutenberg University, Mainz, Germany

Corresponding Author:
Perikles Simon, MD, PhD
Department of Sports Medicine, Disease Prevention and Rehabilitation
Institute of Sports Science
Johannes Gutenberg University
Johann-Joachim-Becher-Weg 31
Mainz, 55099
Germany
Phone: 49 6131 ext 3923586
Fax: 49 613139235
Email: simonpe@uni-mainz.de

Abstract

Background: Physical inactivity is a major risk factor for nonalcoholic fatty liver disease (NAFLD). Exercise-based prevention interventions for improving cardiorespiratory fitness are a recommended complementary treatment for NAFLD. Achievement of minimally effective physical activity to improve cardiorespiratory fitness among patients typically involves high personal and financial expenses in face-to-face settings. We designed an eHealth approach for patients with NAFLD to improve the cardiorespiratory fitness and report the first results of the HELP (Hepatic Inflammation and Physical Performance in Patients With NASH [nonalcoholic steatohepatitis]) study.

Objective: We aimed to assess the effectiveness of an 8-week, tailored, Web-based exercise intervention for cardiorespiratory fitness improvement, expressed as peak oxygen uptake (peak volume of oxygen [VO_{2peak}]), in patients with histologically confirmed NAFLD.

Methods: In a 24-month period, 44 patients were enrolled into an 8-week, prospective, single-arm study with 12 weeks of follow-up. After a medical examination and performance diagnostics, a sports therapist introduced the patients to a Web-based platform for individualized training support. Regular individual patient feedback was provided to systematically adapt the weekly exercise schedule, which allowed us to monitor and ensure patient adherence to strength and endurance training and optimize the step-wise progressive exercise load. Exercise progression was based on an *a priori* algorithm that considered the subjective rate for both perceived exhaustion and general physical discomfort. The VO_{2peak} was assessed at baseline and at the end of the study by spiroergometry.

Results: A total of 43 patients completed the intervention with no adverse events. The VO_{2peak} increased significantly by 2.4 mL/kg/min (8.8%; 95% confidence interval [CI]: 1.48-3.27; P<.001) accompanied by a reduction of 1.0 kg in a body weight (95% CI: 0.33-1.58; P=.004) and 1.3 kg in body fat mass (95% CI: 0.27-2.27; P=.01). In an exploratory analysis, step-wise logistic regression analysis revealed low body fat and VO_{2peak} at baseline and the total minutes of endurance training during the intervention as main contributors to a positive change in VO_{2peak}. Our predictive model indicated that the average patient with NAFLD needed 223 min for stabilization of VO_{2peak} and 628 min for average improvement in VO_{2peak}. However, in patients with a VO_{2peak} approximately 20% higher than the average VO_{2peak}, 628 min were only sufficient to stabilize the VO_{2peak} and >40% reduction in the average fat mass would be required to achieve an average outcome.

Conclusions: This is the first study to show that patients with NAFLD can be effectively supported by a Web-based approach, which can increase the VO_{2peak} to a similar extent as face-to-face interventions. Patients with low body fat and low VO_{2peak} benefited the most from our intervention. In terms of future treatment strategies, NAFLD patients with high body fat may

particularly benefit from body-fat reduction through a strict nutritional intervention, subsequently enabling a more effective exercise intervention.

Trial Registration: ClinicalTrials.gov NCT02526732; https://clinicaltrials.gov/ct2/show/NCT02526732 (Archived by WebCite at http://www.webcitation.org/74pXhXXfq)

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.8607

(J Med Internet Res 2019;21(1):e11250) doi:10.2196/11250

KEYWORDS

exercise; fatty liver; lifestyle; NAFLD; treatment; Web-based

Introduction

Sedentary behavior and an unhealthy diet are common in Western industrialized countries [1,2]. Modern lifestyle increases the risk for chronic diseases such as metabolic syndrome [2-5]. According to the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in adults, metabolic syndrome is defined by the presence of abdominal obesity, hypertriglyceridemia, low high-density lipoprotein cholesterol levels, hypertension, and impaired fasting glycemia [6]. In the last few years, nonalcoholic fatty liver disease (NAFLD) has gained attention, owing to its highest increase in the incidence among chronic liver diseases worldwide [3,4,7-9]. Some researchers consider NAFLD as a hepatic manifestation of metabolic syndrome, whereas others consider it a consequence of metabolic syndrome [3,10-13].

Irrespective of age and ethnicity, 20%-30% of the general population shows fatty changes in the liver [8,11,14-17], and the prevalence of NAFLD is higher in patients with diabetes than in patients without diabetes [18]. NAFLD is a benign, preliminary-stage disease with the potential to progress from simple steatosis to nonalcoholic steatohepatitis (NASH), cirrhosis, and finally, hepatocellular carcinoma [9,19-22]. The pathways driving this progression are numerous and complex, [22,23] and not every patient with NAFLD develops cirrhosis-related complications [9,24]. However, patients with NAFLD have a higher mortality rate than the general population [9,24,25]. Most patients with NAFLD are asymptomatic [16,19,26], but some experience unspecific symptoms such as fatigue [26,27] and depression [28], which additionally affect the health-related quality of life [29,30]. If NAFLD is left untreated, most patients will develop diabetes in the long-term [9]. To improve the condition of the liver and reduce additional risk factors, changes towards a balanced nutrition and a more physically active lifestyle are recommended in daily life [21,26,31-34]. The Practice Guidelines of the American Association for the Study of Liver Diseases recommends a loss of at least 3%-5% of body weight to improve steatosis [35]. The current recommendation for adult patients with NAFLD or NASH is a physical activity target of at least 150 min of moderate-intensity exercise per week or 75 min of vigorous-intensity exercise per week [36]. In addition, strengthening exercises should be performed twice a week [36]. Most people with NAFLD are unaware of the presence of the disease due to the absence of any specific symptoms. Therefore, NAFLD is occasionally self-caused [37] and develops and progresses over years. Studies showed a reduced physical activity level (intensity and amount) in patients with NAFLD compared to healthy individuals [12,27,38] and a suboptimal cardiorespiratory fitness, with <20% of patients meeting the recommended physical activity [39]. In a survey conducted by Kistler et al, 54% of participants were inactive, and 57% of them did not perform any recreational activity; although the remaining 43% performed some activity, it was not enough to achieve the goals recommended [40]. Besides the decreased activity level, prolonged sitting time is associated with a higher prevalence of NAFLD [41,42]. At diagnosis, patients are encouraged to immediately change many aspects of their daily routine, which requires them to overcome different barriers and obstacles; for example, time and place constraints are common hurdles in maintaining regular activity [4,37,43,44]. Changing one's lifestyle is not easy, especially for patients with highly sedentary habits [45]. Consequently, regular motivational support from experts is needed [45]. Thus, advances in modern technologies should be considered for promoting health-conscious behavior [43]. A survey conducted by the Pew Research Center in 2015 showed that 84% of American adults have access to a computer and regularly use the internet [46]. In 2005, 75% of internet users searched for health information, and 42% of them searched for specific information about exercise and training [47]. The possibility to reach and support large numbers of patients via the internet [48] can thus be a cost-effective way to improve and maintain an active lifestyle [49]. Further, frequent issues like time and place constraints for joint exercise could be neglected with an eHealth approach. Web-based interventions with cancer patients were the first to report promising results [50,51]. The aim of our prospective, non-randomized, pilot study was to determine whether online support aids patients with NAFLD or NASH in establishing and maintaining a regular level of physical activity and whether individualized training recommendations improve the overall physical fitness determined by the peak oxygen uptake (peak volume of oxygen [VO_{2peak}]) and body composition.

Methods

Participants

The HELP (Hepatic Inflammation and Physical Performance in Patients With NASH) study is a prospective, single-arm study in patients with histologically confirmed NAFLD and explored the feasibility and effectivity of an individualized exercise intervention. A total of 46 patients were recruited from August 2015 to December 2017. The study was registered at ClinicalTrials.gov (NCT02526732).

Before the study was initiated, a focus group was conducted at the Department of Sports Medicine. The team of experts tested the website over a period of 4 weeks. During this time, the distribution of roles, usability, and communication features were tested and revised. No changes in structure or content were made during the trial period. All study participants were monitored and supported by the same sports therapist. A 24-hour turn-around time was mandated for responses to requests from the participants.

The inclusion criteria were age of 18-70 years and histologically proven NAFLD. Subjects were excluded if they had bariatric operation in the past 5 years; body mass index (BMI) <18.5 kg/m^2 or >45 kg/m^2; instable coronary heart disease; coronary interventions in the past 6 months; stroke in the past 6 months; high-grade coronary artery disease (II-IV); chronic obstructive pulmonary disease; renal or metabolic abnormalities; uncontrolled hypertension; other liver diseases such as hepatitis; decompensated liver cirrhosis; hepatocellular carcinoma; alcohol consumption >30 g/day in men and >20 g/day in women; pregnancy; medications that can cause secondary NASH, such as corticosteroids; other immunological or inflammatory diseases (eg, systemic lupus erythematosus); musculoskeletal disorders; and phenprocoumon therapy.

Study Design

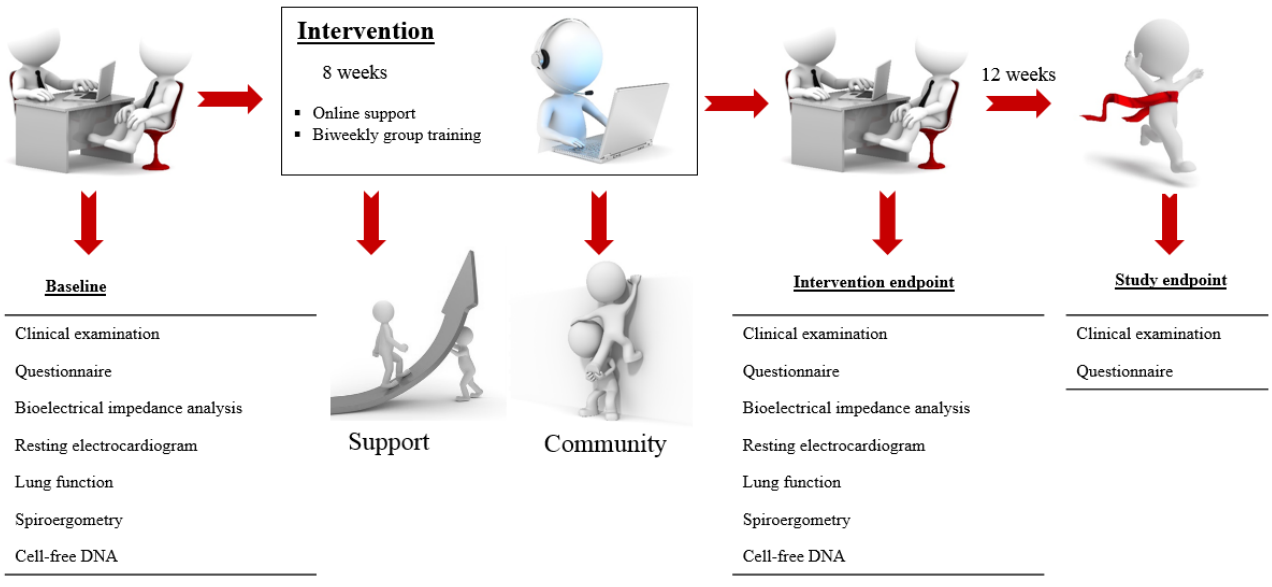
The primary outcome was defined as a change in the VO_{2peak} from the baseline. The secondary outcome measures included changes in body composition, lung function, and changes in performance parameters from the cardiopulmonary exercise test. Assessment of the primary and secondary outcomes was performed before and after the 8-week intervention. Particular attention was paid to the acceptance and safety of the Web-based support concept. From a physical intervention point of view, acceptance of the concept was assessed by the performed exercise recommendations, where completion of 80% of the recommended exercises indicated good adherence. Considering that this was an eHealth approach, acceptance was assessed on the basis of user behavior and consequent weekly feedback. Safety was assessed by the occurrence of overloading or injuries.

Patients underwent physical examinations, laboratory tests, and an ultrasound prior to inclusion in the study. After providing informed consent, the patients performed a cardiopulmonary exercise test until exhaustion at the Department of Sports Medicine, University Mainz. In addition, the body composition was measured using a bio-impedance analyzer (InBody 3.0; Biospace, Seoul, South Korea). A standard 12-lead resting electrocardiogram and a pulmonary function test (spirometry by Body Box 5500; Medisoft, Sorinnes, Belgium) were also performed. Further details about the testing procedure are extensively described elsewhere [52,53]. After the exercise test, each patient was registered and trained by the administrator on the webpage. The registration and explanation process took approximately 1 hour. A detailed explanation, a manual for the homepage, a heart rate monitor, and three elastic tapes were provided to each patient. A detailed illustration on the measurements is available as a video clip in Multimedia Appendix 1. All patients underwent the abovementioned clinical and sports medical examinations at the start of the study and after 8 weeks of the intervention. The clinical examination was additionally performed 12 weeks after the end of the intervention (Figure 1).

Intervention

Patients received tailored exercise recommendations via internal messages on the system on a weekly basis [53]. Depending on the initial exercise test and the subjective feedback from the patients during the intervention period of 8 weeks, the exercise program was adjusted for each patient, considering the American College of Sports Medicine's guidelines for exercise testing and prescription [54]. To avoid early dropouts, moderate-intensity exercise for 3 sessions per week was chosen (endurance training twice [walking or running] and strength training once [major muscle groups]). The program was intensified after a 4-week familiarization to reach a frequency of 5 sessions per week (endurance training thrice and strength training twice) for the remaining 4 weeks.

Figure 1. Study flow chart.



In addition to the frequent interaction with a counselor, peer support is considered a cornerstone in this Web-based concept. Therefore, a discussion board and a chatroom were established to improve social support and adherence [55]. However, due to issues in data-protection regulations and the ethics protocol, activities were not controlled, stored, or analyzed.

Strength Training

A program of 10 strength exercises was carried out in a prescribed sequence to stimulate muscular strength in the major muscle groups. Detailed illustrated instruction and video files for the exercises were available on the website. Training individualization was achieved by varying the number of repetitions or number of sets.

Endurance Training

The individualized endurance training program was based on lactate measurements. The intensity of the jogging program was controlled by a heart rate monitor (FT1; Polar, Kempele, Finland). After an initial continuous method with a heart rate at the lactate threshold, the training followed the interval method (eg, 2 x 4 min, 2 x 3 min, 2 x 2 min with 2-min rest) at a higher heart rate. The intended intensity of training was achieved by adjusting the interval time or adding additional intervals.

Training Progress

Endurance training and strength training, consisting of bodyweight exercises and exercises with elastic tapes, were the main content of the training recommendations. Each training started with a 5-min warm up and was followed by a 5-min cool-down phase. A selection of relaxation and breathing exercises was also available on the website. At the end of each week, the patients sent in a filled schedule with important information such as average heart rate, resting heart rate, and training time. In addition, the following parameters were assessed to allow modification of the training intensity and duration for the following week: (1) The patient's individual assessment of pain and training load using the modified Borg scale (score range, 1-10) [56] after each training session, and (2) the traffic light principle to regulate the intensity of the next week's training. Depending on individual feedback, an increase or decrease in the training recommendations was possible, wherein the pain value was dominant for the decision [53]. The weekly feedback ensured appropriate load according to individual abilities and assessment of compliance. In addition, group training was offered biweekly at the sports center of the University Mainz.

Statistical Procedures

Statistical analysis was performed using SPSS Statistics, version 23 (IBM Corp., Armonk, NY), and JMP (SAS Institute, Cary, NC). Descriptive statistics were used for presenting baseline characteristics and the user behavior of the website. Variables were described using mean, median, and SD. Normal distribution was tested using the Shapiro-Wilk test due to the small sample size. In case of normal distribution, the paired student *t* test was used to determine within-group differences before and after (at 8 weeks) the intervention. Intention-to-treat analysis was performed, and the data were processed according to the last observation carried forward method. A *P* value <.05 was considered statistically significant.

For investigation of factors that contribute to changes in VO_{2peak}, we employed a two-step procedure. We first computed a step-wise feed-forward logistic regression analysis. To ensure stringent inclusion criteria, we fed the model with the baseline data on anthropometrics, including body composition and performance data (Table 1), and data on endurance training (for 8 weeks) and the total exercise time (strength and endurance training in minutes for 8 weeks). Only 3 factors reached a significance level set at .05 for entering a single variable into the regression equation. These factors were then used to compute a logistic regression analysis. Fold-change in VO_{2peak} was normalized by a normalization procedure using the inverse of the squared values, as suggested by Box-Cox analysis.

Results

Baseline Characteristics

A total of 46 patients were screened. After exclusion of 2 patients, 44 patients were finally included in the study (Figure 2). One patient dropped out (2.3%) during the intervention period.

With regard to the weight status, 3 (6.8%) patients had a normal weight status, 15 (34.1%) were overweight, and 26 (59.1%) were obese. All patient characteristics at the baseline are summarized in Table 1. The mean age of all patients was 42 (SD 10.9) years. Of the 44 patients, 41 (93.2%) had a BMI >25 kg/m² and >27% percent of body fat, and the mean BMI of all patients was 31.2 (SD 4.3) kg/m². In total, 1169 exercise recommendations (730 endurance and 439 strength) were performed, and 207 recommended workouts were cancelled due to different reasons.

Primary Outcome

At baseline, the mean VO_{2peak} was 27.2 mL/kg/min. After the intervention, the VO_{2peak} significantly increased by 8.8% (from 27.2 mL/kg/min [SD 5.1] to 29.6 mL/kg/min [SD 5.4]; 95% CI: –3.27 to –1.48; *P*<.001; Figure 3).

We employed logistic regression analysis to assess the combined effects of the variables shown in Table 1 on the fold change in the maximum volume of oxygen (VO_{2max}) by step-wise feed-forward logistic regression analysis. A multiple linear regression model with 3 independent predictors emerged on the basis of 43 total observations (df=3; F=8.03; r^2=0.38; *P*<.001). All predictors had a significant influence and a corrected power of >80% with VO_{2peak} at baseline (t=–3,77; standard effect size=–0.12; 95% CI: –0.18 to –0.05; *P*<.001), total minutes of endurance training during the intervention period (t=3.27; standard effect size=0.09; 95% CI: 0.04-0.15; *P*=.002), and body fat (%) at baseline (t=–3.22; standard effect size=–0.10; 95% CI: –0.17 to –0,04; *P*=.003). The model indicated that participants in the program with average body fat percentage (27.9%) and average VO_{2peak} at baseline (27.1 mL/kg/min) would need approximately 223 minutes of training within the intervention period to maintain their baseline VO_{2peak} (Figure

4a), but 628 minutes of training were required to reach the average improvement of basal VO_{2peak} of approximately 8% (Figure 4b). However, a high VO_{2peak} at baseline with the average fat mass would lead to a significantly lower outcome (Figure 4c) that could, in principle, be compensated by low fat mass with high baseline VO_{2peak} (Figure 4d).

Figure 2. Patient flow chart.

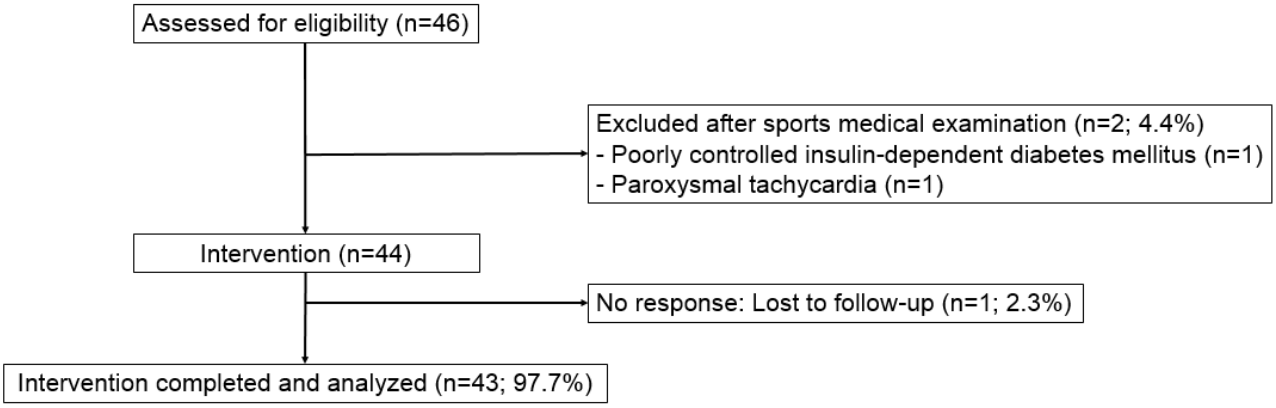


Table 1. Baseline characteristics of patients enrolled in the HELP study (N=44).

Characteristics	Values
Age (years), n (%)	
<30	5 (11.4)
30-60	38 (86.4)
>60	1 (2.3)
Male, n (%)	30 (68.2)
Height (cm), mean (SD)	175 (10.3)
Weight (kg), mean (SD)	95.9 (17.4)
BMI[a] (kg/m²), n (%)	
Overweight (30< BMI >25)	15 (34.1)
Obese (BMI >30)	26 (59.1)
Body composition, mean (SD)	
Body fat (kg)	26.7 (8.2)
Body fat (%)	27.9 (7.4)
Lean body mass (kg)	64.8 (14.1)
Spirometry, mean (SD)	
Forced vital capacity (% norm)	107.5 (13.3)
Forced expiratory volume (% norm)	96.3 (16.3)
Spiroergometry, mean (SD)	
Resting heart rate (bpm)	79 (10.2)
VO_{2peak}[b] (mL/kg/min)	27.2 (5.1)
Watt max	135.1 (42.9)
Watt individual anaerobic threshold	96.1 (21.5)
Borg value max score (range, 6-20)[c], mean (SD)	18.5 (1.5)
Heart frequency max, mean (SD)	172 (16)

[a]BMI: body mass index.

[b]VO_{2peak}: peak volume of oxygen.

[c]The full Borg Scale (score range, 6-20) was used for the cardiopulmonary exercise test.

Figure 3. Individual changes in VO_{2peak} from baseline to end of the study. VO_{2peak} improved by 2.4 mL/kg/min.

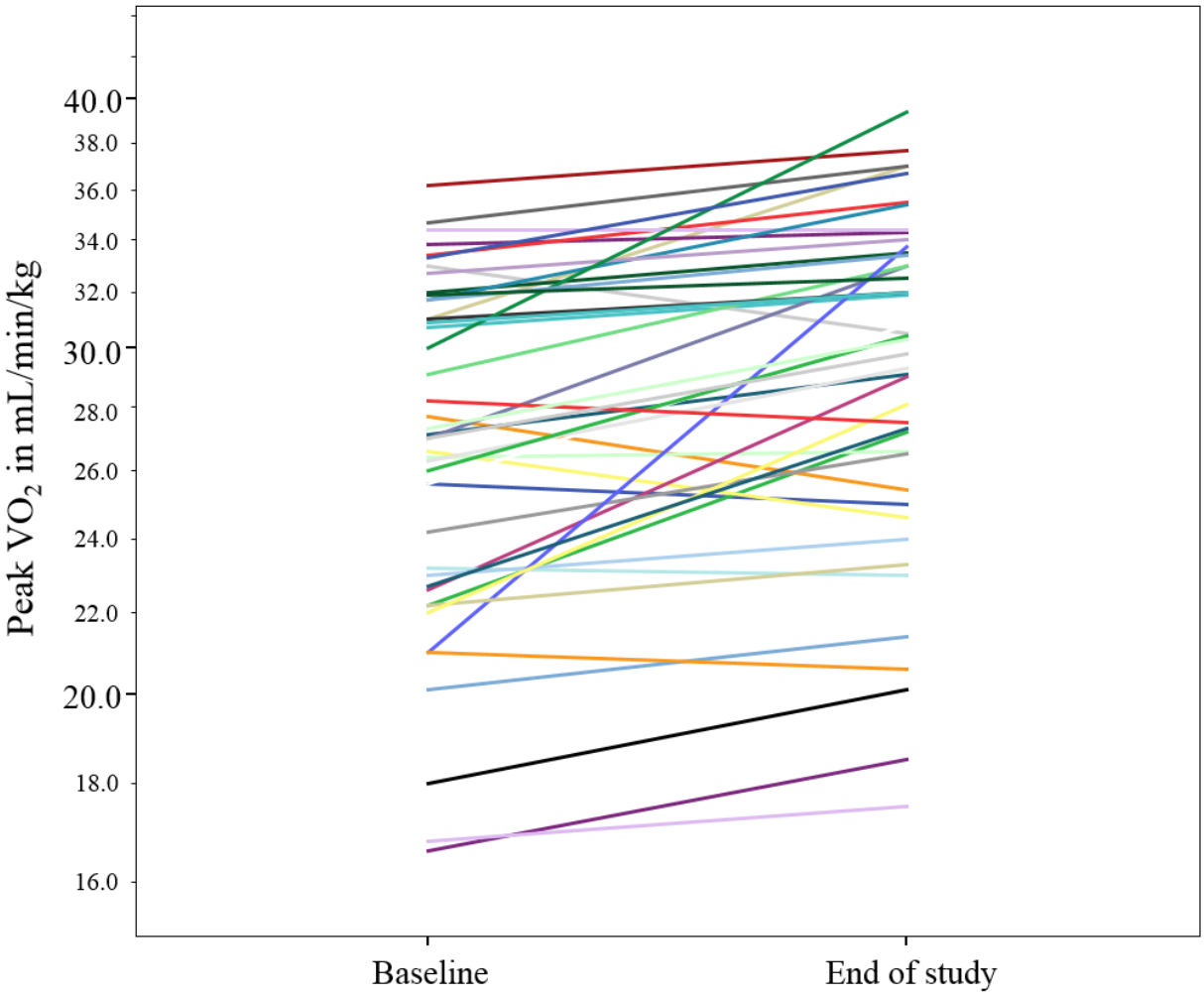
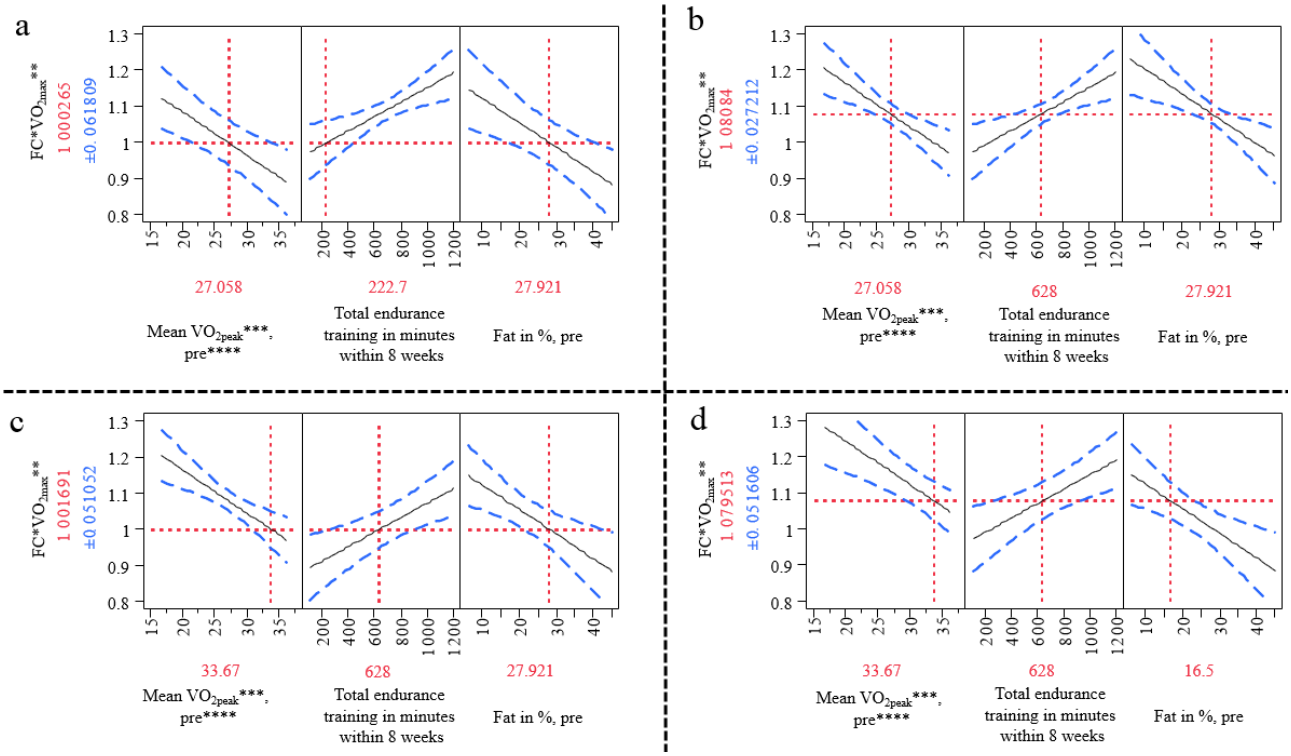


Figure 4. Predictive analysis for fold-change in VO_{2max}. (a) According to this model, 222 min of endurance training is needed to stabilize VO_{2max} in the collective for a person with an average VO_{2peak} of 27 mL/kg/min and average body fat of 27.9%. (b) For an improvement of approximately 8% VO_{2max}, an endurance training load of at least 600 min over 8 weeks is necessary. (c) A higher initial VO_{2peak} leads to a reduced effect of the 628 min of endurance training within 8 weeks on the primary outcome, VO_{2max}. (d) In principal, lower body fat (%) could compensate for a higher VO_{2peak} at baseline (33.67 mL/kg/min) and yield an 8% improvement in VO_{2max} with the same training load. Solid black regression lines indicate the linear effect of VO_{2peak} at the start of the study, total minutes of endurance training during the intervention period of 8 weeks, and body fat percentage at the start of the study. Dashed blue lines indicate the respective upper and lower 95% confidence intervals for the regressions. *FC: Fold-change. **VO_{2max}: maximum volume of oxygen. ***VO_{2peak}: peak volume of oxygen. ****Pre: before the study.



Secondary Outcomes

Significant changes were observed in body weight and BMI (95% CI: 0.33-1.58; *P*=.004 and 95% CI: 0.14-0.54; *P*=.001, respectively). With regard to the body composition, there was a significant reduction in body fat (95% CI: 0.27-2.27; *P*=.01); a consequent reduction in the percentage of body fat (95% CI: 0.26-2.11; *P*=.01); and a slight, but not significant, increase in lean body mass (95% CI: –1.39 to 0.46; *P*=.31; Table 2). There was a trend towards a low resting heart rate (95% CI: –0.18 to 7.22; *P*=.06), but this result was not statistically significant. Further, no changes in lung function, expressed as forced expiratory volume (95% CI: –4.11 to 1.38; *P*=.32) and vital capacity (95% CI: –1.09 to 3.41; *P*=.31), were observed.

Significant changes in the power, expressed as Watt max (95% CI: –18.46 to –10.17; *P*<.001) and Watt at the individual anaerobic threshold (95% CI: –7.00 to –2.05; *P*=.001) were observed (Figure 5). The maximum heart frequency remained unchanged, and the subjective perception of exhaustion, expressed as the Borg value, decreased significantly from baseline to the end point of the study (95% CI: –0.05 to 0.95; *P*=.02; Table 2).

Acceptance of the Program From an EHealth Perspective

During the intervention period, regular communication and feedback were easily achieved using the webpage. In some cases, the patients did not send the exercise feedback on time. Therefore, 2.8 (SD 3.8) reminders were sent to the participants to ask for the exercise feedback. The user behavior in terms of log-in duration and frequency is presented in Table 3. The participants' average length of a visit was approximately 12 min, and the average login frequency was 13 times during the intervention period.

The typical expected attrition in registration frequency and duration [57] was observed in this trial (Figures 6 and 7). Nevertheless, a timely response was achieved, even in patients who did not continue to use the webpage, by interacting via conventional email (Table 3).

Table 2. Study results (N=44).

Characteristics	Baseline, mean (SD)	After the study, mean (SD)	Difference (%)	*P* value
Weight (kg)	95.9 (17.4)	95.0 (17.8)	0.9 (0.9)	0.004
BMI[a] (kg/m²)	31.2 (4.3)	30.8 (4.4)	0.4 (1.3)	0.001
Overweight (30< BMI >25)	15 (34.1)[b]	14 (32.6)[b]	1 (6.7)	—[c]
Obese (BMI >30)	26 (59.1)[b]	26 (59.1)[b]	0 (0)	—
Body composition				
Body fat (kg)	26.7 (8.2)	25.5 (9.0)	1.2 (4.5)	0.01
Body fat (%)	27.9 (7.4)	26.8 (8.4)	1.1 (3.9)	0.01
Lean body mass (kg)	64.8 (14.1)	65.2 (14.2)	0.4 (0.6)	0.31
Spirometry				
Forced vital capacity (% norm)	107.5 (13.3)	106.3 (14.2)	1.2 (1.1)	0.31
Forced expiratory volume (% norm)	96.3 (16.3)	97.6 (13.3)	1.3 (1.3)	0.32
Spiroergometry				
Resting heart rate (bpm)	79 (10.2)	75 (11.5)	4 (5)	0.06
VO_{2peak}[d] (mL/kg/min)	27.2 (5.1)	29.6 (5.4)	2.4 (8.8)	<0.001
Watt max	135.1 (42.9)	149.5 (49.5)	14.4 (10.7)	<0.001
Watt individual anaerobic threshold	96.1 (21.5)	100.6 (24.6)	4.5 (4.7)	0.001
Borg value max (range, 6-20)[e]	18.5 (1.5)	18 (1.8)	0.5 (2.7)	0.03
Heart frequency (max)	172 (16)	172 (14.8)	0 (0)	—

[a]BMI: body mass index.

[b]Values are n (%) rather than mean (SD).

[c]Not applicable.

[d]VO_{2peak}: peak volume of oxygen.

[e]The full Borg Scale (score range, 1-10) was used for the cardiopulmonary exercise test.

Figure 5. Fold changes for maximum Watt, maximum volume of oxygen (VO_{2max}, body fat, and Watt at the individual anaerobic threshold (IAT). Error bars indicate 95% confidence intervals.

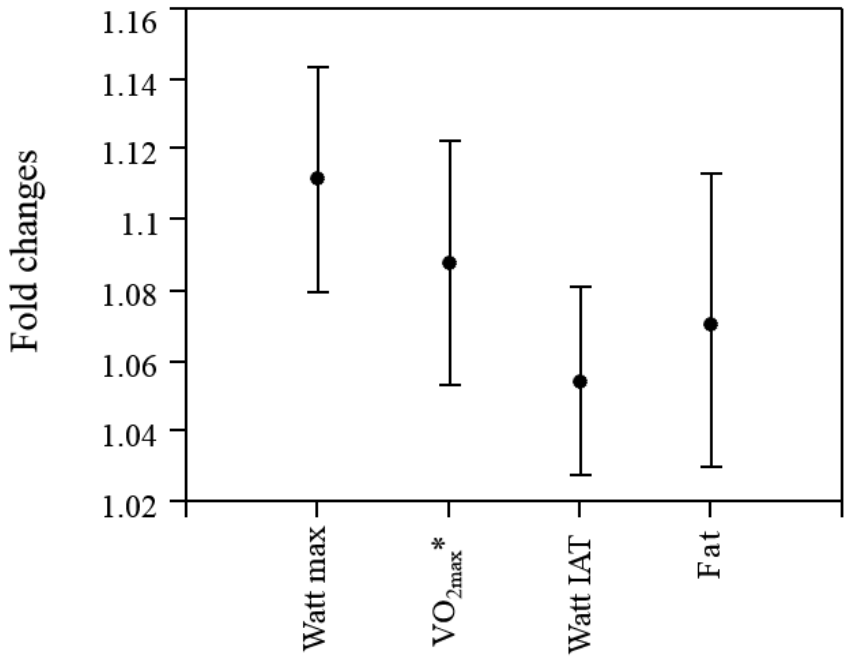


Table 3. User behavior for the homepage during the intervention period (N=43).

Characteristics	Sum	Mean (SD)	Range
Total number of logins	557	13.0 (8.0)	3-38
Total login duration (min)	6548	152.3 (93.8)	16-367
Total number of reminders	120	2.8 (3.8)	0-18
Use of email instead of the website for exercise feedback, n	165	3.8 (3.7)	0-8

Figure 6. Login frequency during the intervention period of 8 weeks. Error bars indicate 95% confidence intervals.

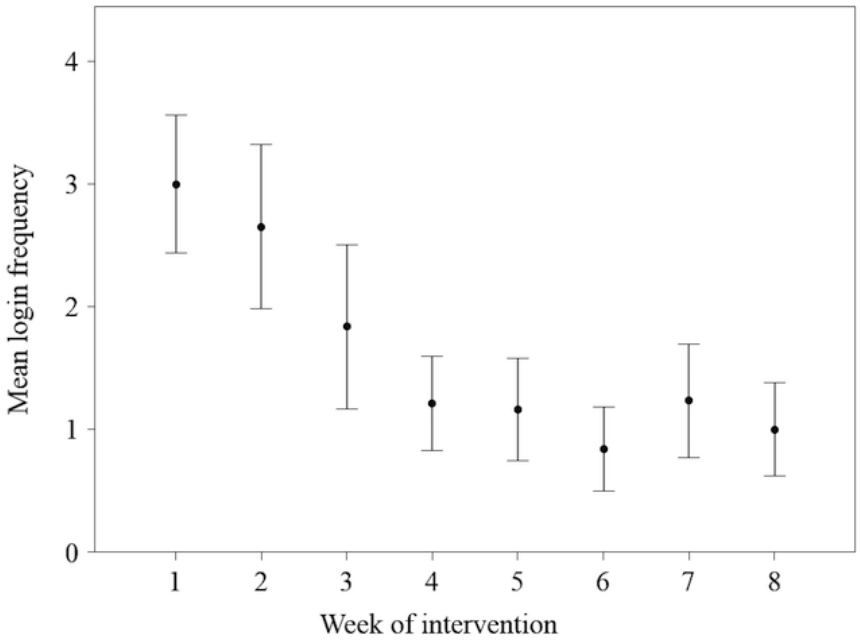


Figure 7. Login duration in minutes during the intervention period of 8 weeks. Error bars indicate 95% confidence intervals.

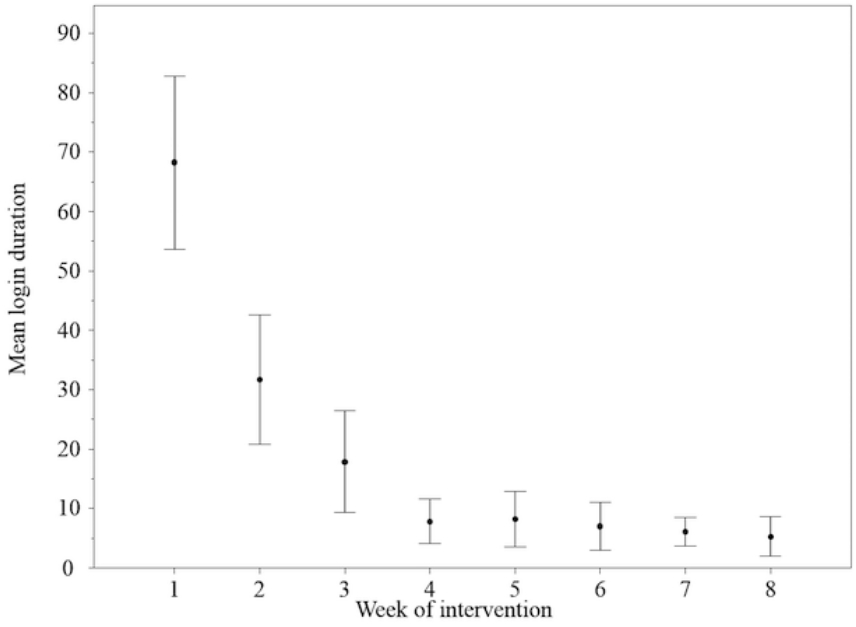


Figure 8. Development of the physical activity level, expressed as weekly endurance training in minutes. The dashed line indicates the recommended vigorous activity goal of 75 min per week by the World Health Organization. Error bars indicate 95% confidence intervals.

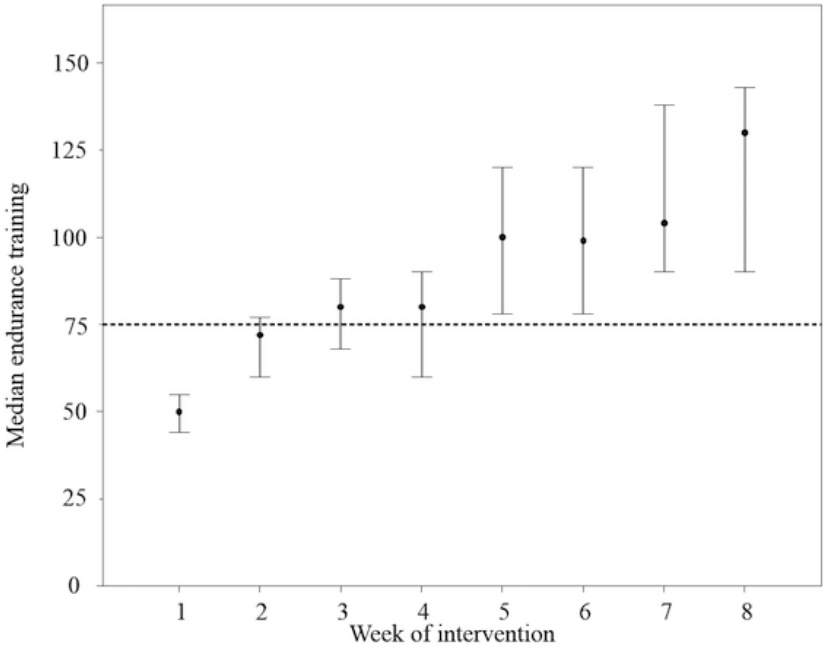


Table 4. Exercise profile (N=43).

Characteristics	Sum	Mean (SD)	Range
Total physical activity within 8 weeks (min)	52373	1218 (330.0)	568-1801
Endurance training within 8 weeks (min)	29104	677 (220.2)	173-1118
Interruption of exercise training, n	207	4.81 (5.1)	0-17
Additional workouts, n	72	1.7 (2.8)	0-13

Acceptance of the Program From a Physical Intervention Point of View

The online exercise concept was well accepted by the patients. The physical activity level increased steadily over the period of 8 weeks (Figure 8). After an initial familiarization period, the exercise recommendations increased progressively. In the second half of the intervention, the patients reached and exceeded the activity recommended by the World Health Organization [58]. The study participants performed 72 additional workouts (eg, hiking or playing volleyball or badminton; Table 4). However, the participants were not obliged to record other leisure time activities, and therefore, the additional exercises were not further examined. The adherence to the Web-based exercise concept, expressed as ≥80% of the endurance workouts, was good: 32 participants (74%) performed ≥80% of the recommended endurance workouts. Training interruptions were either documented in the weekly exercise schedule or in reply to a request from the sports therapist. Common reasons for breaks were deadlines (eg, congress participation or workshops), medical reasons (eg, cold, inflammation, blisters, headache, or food poisoning), external conditions (eg, high temperature or heavy rain), or private reasons.

Safety

The intervention in this group of patients with liver disease was safe. No serious adverse events occurred during the study period. Our traffic light principle for exercise modification protected the participants against overload. The weekly communication between the participant and the sports therapist was essential for the tailored feedback [53]. Neither the patients nor the attending physician reported injuries, physical overload, or serious concerns about the health status during the intervention period. Notably, due to physical discomfort, 2 patients had to be assigned to bicycle exercise instead of the prescribed exercise.

Discussion

Overview

Current guidelines recommend lifestyle changes as the primary approach for treating obesity and NAFLD; however, only a few studies have focused on physical activity, and neither the type nor the intensity of appropriate exercise has been defined for NAFLD. In addition, it has been suggested that physical deconditioning of patients with NAFLD leads to the inability to adhere to exercise recommendations [12,38,39]. The HELP study explored the feasibility and efficiency of a Web-based and patient-centered exercise support concept.

The Role of Weight Loss

Weight loss is a major goal in the treatment of NAFLD [59]. The guidelines and experts in the field recommend a weight loss of at least 3%-5% to improve steatosis [35,60]. However, there are critical limits to the weight loss [61]: A weight loss of 1.6 kg per week should not be exceeded, as it could potentially provoke portal inflammation or portal fibrosis [60,61]. In our study, 3 patients showed a normal weight status; therefore, weight reduction was not needed. The term *weight management* is more accurate than weight loss in this context. We found a significant, but extremely low, weight change among the patients, which is in accordance with the findings of other exercise studies [62-64]. Weight gain is a result of a high energy intake and low energy expenditure [65]. Weight reduction is only possible if the energy expenditure persistently exceeds the energy intake [31]. Therefore, diet is a necessary aspect of weight reduction. Furthermore, our prediction model showed that a reduction in the fat mass percentage is crucial to significantly improve cardiorespiratory fitness with achievable physical activity. Regular exercise supports energy expenditure, but conscious nutrition is essential for control of energy intake. Moreover, the absence of weight loss might partly be explained by a moderate shift from fat mass to fat-free mass. Regular activity reduces body fat and increases lean body mass. Another explanation for not observing weight loss during the intervention is insufficient negative energy balance due to a low starting intensity [66]. For effective weight loss, a longer duration and increased intensity of exercise is required. Nevertheless, exercise studies show promising results in terms of decreased insulin and homeostasis model assessment index [63,67], improved cardiorespiratory fitness [62,68], reduced hepatic and visceral lipid levels [63,69], reduced liver enzyme levels, and modulated liver fat content [34], irrespective of the nutrition intake. Exercise combined with an adjusted diet shows strong effects on weight change, but exercise also has independent modes of action. Thus, physical activity or structured exercise recommendations should be strongly promoted due to their additional benefits in the absence of weight loss.

Fitness Improvement

Takahashi et al (2015) assessed the efficiency and safety of two simple resistance exercises in 53 patients with NAFLD [67]. After a 12-week period, patients in the intervention group had a significantly increased mean fat-free mass (–0.24 [SD 0.88] vs 0.30 [SD 0.67] kg; P=.01) and muscle (–0.24 [SD 0.82] vs 0.25 [SD 0.70] kg; P=.02) compared to the control group [67], which are in line with our findings. Another previous study reported a change of nearly 9% in VO_{2peak} when participants in the intervention group trained 5 times per week for 16 weeks [70]. The subjects exercised under supervision once a week and were encouraged to perform the remaining 4 sessions at home. The endurance training was controlled by heart-rate measurement [70]. In our study, a combination of the heart rate, as an objective measure, and the Borg value, as a subjective measure, were used to determine the intensity. Borg values are considered an appropriate measure for monitoring and regulating exercise intensity [34,71,72]. The results of this study, with a short intervention period of 8 weeks, are consistent with those of recent face-to-face research studies with respect to changes in body composition and VO_{2peak}, demonstrating the efficacy of our eHealth approach [34,63,64,67,70,73]. For instance, in a study by Keating et al (2015), the participants were supervised and monitored using Borg values and heart rate during the intervention period of 8 weeks [64]. Depending on the group allocation, the participants were instructed to perform 2 or 3 sessions of aerobic exercise training a week; the exercise quantity is comparable to that used in our study. With regard to the exercise quality, Keating et al (2015) compared 3 different training approaches in order to determine whether intensity or volume is more effective [64]. A total of 47 obese adults trained for 8 weeks with low-intensity and high-volume exercise, high-intensity and low-volume exercise, or low-intensity and low-volume exercise (or were prescribed a stretching and self-massage program as placebo); all intervention groups showed similar changes in VO_{2peak} (by 2.29 [SD 0.77] mL/kg/min, P<.01; 2.99 [SD 0.48] mL/kg/min, P<.01; and 2.24 [SD 0.54] mL/kg/min, P<.01, respectively), which is comparable to our results. Therefore, the investigators concluded that volume and intensity were both efficient [64]. However, with respect to adherence, another study reported that high exercise frequency is better accepted than high exercise intensity [74]. Furthermore, high exercise intensity resulted in a higher percentage of exercise-related injuries [74]. Despite the positive effects of regular physical activity, independent of intensity and volume, the most important challenge is adherence to exercise [75].

The Importance of Regular Support

Physical activity for 5 days per week is recommended for the general population [58]. In addition, resistance training should be performed at least 2 days a week. In contrast to the findings of Perri et al [74], many study participants in this trial reported that they struggled with integration of the demanding volume in the second half of the intervention. Berzigotti et al (2016) summarized the beneficial effects of exercise on the health of patients with NAFLD, but also reported high dropout rates in physical activity trials among these patients [3]. There is an urgent need to counteract the sedentary habits of patients with NAFLD. Despite the proven positive effects of regular physical activity and a healthy diet, many people lack long-term motivation [76]. Engaging in less physical activity increases the risk of fatty changes in the liver [41,77,78]. In a previous study, the physically inactive group showed a significantly higher prevalence of fatty liver changes [78]. This finding is supported by the results of a study by Perseghin et al (2007), who showed an association between habitual physical activity and intrahepatic fat content [79]. Furthermore, a large cross-sectional study by Ryu et al (2015) fully supported this finding: They showed that prolonged sitting times are positively associated with the prevalence of NAFLD [42]. Therefore, supporting patients to achieve and sustain regular activity is a key issue in NAFLD management [12]. Due to the pronounced sedentary lifestyle, starting with a low training volume and intensity is recommended for motivational reasons. Self-chosen sitting times should be reduced, and barriers for regular exercise should be identified and eliminated [80]. Intensive-exercise interventions carried out under supervision in hospitals or fitness centers [32,34,68,70] impose an unnatural lifestyle on patients for a short period of time. After the intervention, patients quickly

lose motivation and fall back into their old habits [45]. There is a strong need and a high potential for Web-based intervention designs [81,82]. Web-based interventions are essential to bridge the treatment gap between demand and supply [83,84]. In a previous study, patients with diabetes reported that they need a better strategy for transition to subsequent postintervention conditions of less support [85]. Furthermore, from the patient's point of view, scheduling flexibility and geographical proximity are important factors, which should be taken into account [85]. Even the most-powerful individualized exercise program enhances the patient's situation sustainably only if patients are able to adopt the regular activity in their daily routine [86]. Therefore, the main focus is to incorporate an exercise program into the daily routine of patients with NAFLD to promote long-term changes [87] and reverse sedentariness [88]. Regular feedback from a counselor seems to be an important aspect for patients to stay motivated [89,90]. Furthermore, it is important to integrate the patient in the decision-making process [86]. Thus, a change from compliance (implementation of prescription) to adherence (mutual agreement between patient and caregiver) should be achieved [82,86]. Growing interest in utilization of modern technologies such as computers and smartphones could encourage the use feasible ways to improve knowledge and care in a home-based setting. Web-based support allows a flexible scheduling of training and transfer of proper guidance through regular counseling and tailored feedback. In this study, common obstacles such as time constraints (eg, shift work or family responsibilities) or no access to a fitness center (eg, distance, high costs, or lack of sound advice) [37,43,45] could be circumvented by using additional training equipment like pulse watches and elastic bands. Furthermore, a regular, close communication with a team of experts was achieved by communication via the specially designed website. This approach reduced the aforementioned geographical and time-related barriers. In contrast to print-based interventions and face-to-face counseling, Web-based communication reduces costs and has a higher potential to reach a wide range of target groups with tailored support [43,49,91]. Individualized recommendations based on heart frequency and personal feedback, expressed as rating of perceived exertion, seem to be useful in this context [34].

This trial tests the feasibility of a Web-based exercise program in a group of patients with liver disease. In line with the recommendations of guidelines, this is also one of the first investigations to support patients with NAFLD by providing a combined endurance and strength training program.

Strengths and Limitations

This study has several strengths and limitations. The strengths of this study include the performance of a liver biopsy, which is the gold standard for diagnosing liver diseases [7,92]. Furthermore, this is the first Web-based approach for a tailored exercise intervention in this patient group. In contrast to other studies, we had only 1 (2.3%) dropout; our study showed good adherence and tolerance in comparison to a recent review that reported dropout rates of 6%-45% with a similar intervention [75]. Possible advantages of the Web-based approach are the flexible design and exercise implementation in the home environment. This study had a few limitations. First, investigators had to rely, at least partially, on subjective feedback for training adherence without the possibility of visual control by the sports physician. Data on training adherence might be prone to social desirability bias as well as over- or underestimation. Studies have previously discussed the issue of over- or underestimation in home-based training settings [93,94]. To reduce such potential bias, we combined the subjective feedback (Borg rating for training sessions) with an objective measurement (average heart rate). An individual who tended to respond in a socially desirable fashion would need at least some in-depth knowledge in exercise physiology to trick the therapist or would most likely submit nonplausible data. Second, this study lacked a control group. Because of different comorbidities, some patients had to be excluded. Therefore, we do not present a typical NAFLD collective here. Third, missing features and confusing page layout of the website could have affected user behavior. Most patients demanded regular nutritional advice. Some of the participants stated that they neglected their common eating habits due to diverse changes during the intervention period (eg, marriage, job loss, or change of shift). These changes could explain minor changes in the weight status. Finally, the length of the intervention period (8 weeks) was probably too short to show further improvements in weight status or cardiorespiratory fitness.

Conclusions

The present study indicates that 8 weeks of a Web-based, highly individualized, supervised training is safe and feasible for patients with NAFLD. The program significantly improved the VO_{2peak} and body composition. To influence the risk factor sedentariness sustainably and enable a long-term lifestyle change, an exercise program that can be integrated into everyday life is needed. Web-based communication as a connection between the patient and caregiver might be a useful and cost-effective monitoring tool. Close contact with the supervisor can immediately reduce sport-related doubts and anxieties as well as motivational barriers. The Web-based design is the first step to a new way of delivering services to patients with NAFLD and potentially, other diseases. Future studies are required to determine whether regular interaction between the patient and the study team can be maintained in long-term. Additionally, the intervention program presented here could be further supplemented with individualized, nutritional advice by a dietician to further improve the weight status. Finally, an expert should rework the page design and integrate missing features for a more pleasant experience with the webpage.

Acknowledgments

We thank all the participants of this study and the staff of the I. Department of Medicine, University Medical Center and the Department of Sports Medicine, Rehabilitation and Disease Prevention at the Johannes Gutenberg University, Mainz, Germany, for recruiting, sample processing, and support.

Authors' Contributions

PS, JMS, YH, and DP developed the individual study concepts. DP designed the website. PS and DP designed the exercise components, and YH and JMS revised the manuscript. All authors read and approved the final document.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Video clip of the Web-based exercise support concept.

[MP4 File (MP4 Video), 113MB - jmir_v21i1e11250_app1.mp4]

References

1. Finelli C, Tarantino G. Have guidelines addressing physical activity been established in nonalcoholic fatty liver disease? World J Gastroenterol 2012 Dec 14;18(46):6790-6800 [FREE Full text] [doi: 10.3748/wjg.v18.i46.6790] [Medline: 23239917]
2. Gonçalves IO, Oliveira PJ, Ascensão A, Magalhães J. Exercise as a therapeutic tool to prevent mitochondrial degeneration in nonalcoholic steatohepatitis. Eur J Clin Invest 2013 Nov;43(11):1184-1194. [doi: 10.1111/eci.12146] [Medline: 24033085]
3. Berzigotti A, Saran U, Dufour J. Physical activity and liver diseases. Hepatology 2016 Mar;63(3):1026-1040. [doi: 10.1002/hep.28132] [Medline: 26313307]
4. Whitsett M, VanWagner LB. Physical activity as a treatment of non-alcoholic fatty liver disease: A systematic review. World J Hepatol 2015 Aug 08;7(16):2041-2052 [FREE Full text] [doi: 10.4254/wjh.v7.i16.2041] [Medline: 26261693]
5. Booth FW, Roberts CK, Laye MJ. Lack of exercise is a major cause of chronic diseases. Compr Physiol 2012 Apr;2(2):1143-1211 [FREE Full text] [doi: 10.1002/cphy.c110025] [Medline: 23798298]
6. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). JAMA 2001 May 16;285(19):2486-2497. [Medline: 11368702]
7. Adams LA, Angulo P. Treatment of non-alcoholic fatty liver disease. Postgrad Med J 2006 May;82(967):315-322 [FREE Full text] [doi: 10.1136/pgmj.2005.042200] [Medline: 16679470]
8. Masarone M, Federico A, Abenavoli L, Loguercio C, Persico M. Non alcoholic fatty liver: epidemiology and natural history. Rev Recent Clin Trials 2014;9(3):126-133. [Medline: 25514916]
9. Ekstedt M, Franzén LE, Mathiesen UL, Thorelius L, Holmqvist M, Bodemar G, et al. Long-term follow-up of patients with NAFLD and elevated liver enzymes. Hepatology 2006 Oct;44(4):865-873. [doi: 10.1002/hep.21327] [Medline: 17006923]
10. Wilkins T, Tadkod A, Hepburn I, Schade RR. Nonalcoholic fatty liver disease: diagnosis and management. Am Fam Physician 2013 Jul 01;88(1):35-42 [FREE Full text] [Medline: 23939604]
11. Bellentani S, Scaglioni F, Marino M, Bedogni G. Epidemiology of non-alcoholic fatty liver disease. Dig Dis 2010 Dec;28(1):155-161. [doi: 10.1159/000282080] [Medline: 20460905]
12. Hallsworth K, Thoma C, Moore S, Ploetz T, Anstee QM, Taylor R, et al. Non-alcoholic fatty liver disease is associated with higher levels of measured sedentary behaviour and lower levels of physical activity than matched healthy controls. Frontline Gastroenterol 2015 Jan;6(1):44-51 [FREE Full text] [doi: 10.1136/flgastro-2014-100432] [Medline: 25580206]
13. Kotronen A, Yki-Järvinen H. Fatty liver: a novel component of the metabolic syndrome. Arterioscler Thromb Vasc Biol 2008 Jan;28(1):27-38 [FREE Full text] [doi: 10.1161/ATVBAHA.107.147538] [Medline: 17690317]
14. Wong VW, Chu WC, Wong GL, Chan RS, Chim AM, Ong A, et al. Prevalence of non-alcoholic fatty liver disease and advanced fibrosis in Hong Kong Chinese: a population study using proton-magnetic resonance spectroscopy and transient elastography. Gut 2012 Mar;61(3):409-415. [doi: 10.1136/gutjnl-2011-300342] [Medline: 21846782]
15. Browning JD, Szczepaniak LS, Dobbins R, Nuremberg P, Horton JD, Cohen JC, et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. Hepatology 2004 Dec;40(6):1387-1395. [doi: 10.1002/hep.20466] [Medline: 15565570]
16. Satapathy SK, Sanyal AJ. Epidemiology and Natural History of Nonalcoholic Fatty Liver Disease. Semin Liver Dis 2015 Aug;35(3):221-235. [doi: 10.1055/s-0035-1562943] [Medline: 26378640]

17. Koehler EM, Schouten JNL, Hansen BE, van RFJA, Hofman A, Stricker BH, et al. Prevalence and risk factors of non-alcoholic fatty liver disease in the elderly: results from the Rotterdam study. J Hepatol 2012 Dec;57(6):1305-1311. [doi: 10.1016/j.jhep.2012.07.028] [Medline: 22871499]
18. Blachier M, Leleu H, Peck-Radosavljevic M, Valla D, Roudot-Thoraval F. The burden of liver disease in Europe: a review of available epidemiological data. J Hepatol 2013 Mar;58(3):593-608 [FREE Full text] [doi: 10.1016/j.jhep.2012.12.005] [Medline: 23419824]
19. Angulo P, Lindor KD. Non-alcoholic fatty liver disease. J Gastroenterol Hepatol 2002 Feb;17 Suppl:S186-S190. [Medline: 12000605]
20. Starley BQ, Calcagno CJ, Harrison SA. Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection. Hepatology 2010 May;51(5):1820-1832. [doi: 10.1002/hep.23594] [Medline: 20432259]
21. Nobili V, Carter-Kent C, Feldstein AE. The role of lifestyle changes in the management of chronic liver disease. BMC Med 2011 Jun 06;9:70 [FREE Full text] [doi: 10.1186/1741-7015-9-70] [Medline: 21645344]
22. Berlanga A, Guiu-Jurado E, Porras JA, Auguet T. Molecular pathways in non-alcoholic fatty liver disease. Clin Exp Gastroenterol 2014;7:221-239 [FREE Full text] [doi: 10.2147/CEG.S62831] [Medline: 25045276]
23. Ferreira DMS, Castro RE, Machado MV, Evangelista T, Silvestre A, Costa A, et al. Apoptosis and insulin resistance in liver and peripheral tissues of morbidly obese patients is associated with different stages of non-alcoholic fatty liver disease. Diabetologia 2011 Jul;54(7):1788-1798. [doi: 10.1007/s00125-011-2130-8] [Medline: 21455726]
24. Adams LA, Lymp JF, St SJ, Sanderson SO, Lindor KD, Feldstein A, et al. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. Gastroenterology 2005 Jul;129(1):113-121. [Medline: 16012941]
25. Ong JP, Pitts A, Younossi ZM. Increased overall mortality and liver-related mortality in non-alcoholic fatty liver disease. J Hepatol 2008 Oct;49(4):608-612. [doi: 10.1016/j.jhep.2008.06.018] [Medline: 18682312]
26. Rinella ME. Nonalcoholic fatty liver disease: a systematic review. JAMA 2015 Jun 09;313(22):2263-2273. [doi: 10.1001/jama.2015.5370] [Medline: 26057287]
27. Newton JL, Jones DEJ, Henderson E, Kane L, Wilton K, Burt AD, et al. Fatigue in non-alcoholic fatty liver disease (NAFLD) is significant and associates with inactivity and excessive daytime sleepiness but not with liver disease severity or insulin resistance. Gut 2008 Jun;57(6):807-813. [doi: 10.1136/gut.2007.139303] [Medline: 18270241]
28. Tomeno W, Kawashima K, Yoneda M, Saito S, Ogawa Y, Honda Y, et al. Non-alcoholic fatty liver disease comorbid with major depressive disorder: The pathological features and poor therapeutic efficacy. J Gastroenterol Hepatol 2015 Jun;30(6):1009-1014. [doi: 10.1111/jgh.12897] [Medline: 25619308]
29. Younossi ZM, Boparai N, Price LL, Kiwi ML, McCormick M, Guyatt G. Health-related quality of life in chronic liver disease: the impact of type and severity of disease. Am J Gastroenterol 2001 Jul;96(7):2199-2205. [doi: 10.1111/j.1572-0241.2001.03956.x] [Medline: 11467653]
30. Alt Y, Grimm A, Schlegel L, Grambihler A, Kittner JM, Wiltink J, et al. The Impact of Liver Cell Injury on Health-Related Quality of Life in Patients with Chronic Liver Disease. PLoS One 2016;11(3):e0151200 [FREE Full text] [doi: 10.1371/journal.pone.0151200] [Medline: 26990427]
31. Al-Jiffri O, Al-Sharif FM, Abd El-Kader SM, Ashmawy EM. Weight reduction improves markers of hepatic function and insulin resistance in type-2 diabetic patients with non-alcoholic fatty liver. Afr Health Sci 2013 Sep;13(3):667-672 [FREE Full text] [doi: 10.4314/ahs.v13i3.21] [Medline: 24250305]
32. Bacchi E, Negri C, Targher G, Faccioli N, Lanza M, Zoppini G, et al. Both resistance training and aerobic training reduce hepatic fat content in type 2 diabetic subjects with nonalcoholic fatty liver disease (the RAED2 Randomized Trial). Hepatology 2013 Oct;58(4):1287-1295. [doi: 10.1002/hep.26393] [Medline: 23504926]
33. Hannah WN, Harrison SA. Lifestyle and Dietary Interventions in the Management of Nonalcoholic Fatty Liver Disease. Dig Dis Sci 2016 May;61(5):1365-1374. [doi: 10.1007/s10620-016-4153-y] [Medline: 27052013]
34. Hallsworth K, Thoma C, Hollingsworth KG, Cassidy S, Anstee QM, Day CP, et al. Modified high-intensity interval training reduces liver fat and improves cardiac function in non-alcoholic fatty liver disease: a randomized controlled trial. Clin Sci (Lond) 2015 Dec;129(12):1097-1105. [doi: 10.1042/CS20150308] [Medline: 26265792]
35. Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice Guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. Hepatology 2012 Jun;55(6):2005-2023. [doi: 10.1002/hep.25762] [Medline: 22488764]
36. Ratziu V, Bellentani S, Cortez-Pinto H, Day C, Marchesini G. A position statement on NAFLD/NASH based on the EASL 2009 special conference. J Hepatol 2010 Aug;53(2):372-384 [FREE Full text] [doi: 10.1016/j.jhep.2010.04.008] [Medline: 20494470]
37. 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] [Medline: 14769621]
38. Gerber L, Otgonsuren M, Mishra A, Escheik C, Birerdinc A, Stepanova M, et al. Non-alcoholic fatty liver disease (NAFLD) is associated with low level of physical activity: a population-based study. Aliment Pharmacol Ther 2012 Oct;36(8):772-781 [FREE Full text] [doi: 10.1111/apt.12038] [Medline: 22958053]

39. Krasnoff JB, Painter PL, Wallace JP, Bass NM, Merriman RB. Health-related fitness and physical activity in patients with nonalcoholic fatty liver disease. Hepatology 2008 Apr;47(4):1158-1166 [FREE Full text] [doi: 10.1002/hep.22137] [Medline: 18266250]
40. Kistler KD, Brunt EM, Clark JM, Diehl AM, Sallis JF, Schwimmer JB, NASH CRN Research Group. Physical activity recommendations, exercise intensity, and histological severity of nonalcoholic fatty liver disease. Am J Gastroenterol 2011 Mar;106(3):460-8; quiz 469 [FREE Full text] [doi: 10.1038/ajg.2010.488] [Medline: 21206486]
41. Wei H, Qu H, Wang H, Deng H. Associations between sitting time and non-alcoholic fatty liver diseases in Chinese male workers: a cross-sectional study. BMJ Open 2016 Dec 08;6(9):e011939 [FREE Full text] [doi: 10.1136/bmjopen-2016-011939] [Medline: 27609847]
42. Ryu S, Chang Y, Jung H, Yun KE, Kwon M, Choi Y, et al. Relationship of sitting time and physical activity with non-alcoholic fatty liver disease. J Hepatol 2015 Nov;63(5):1229-1237. [doi: 10.1016/j.jhep.2015.07.010] [Medline: 26385766]
43. Napolitano MA, Fotheringham M, Tate D, Sciamanna C, Leslie E, Owen N, et al. Evaluation of an internet-based physical activity intervention: a preliminary investigation. Ann Behav Med 2003;25(2):92-99. [Medline: 12704010]
44. Thomas N, Alder E, Leese GP. Barriers to physical activity in patients with diabetes. Postgrad Med J 2004 May;80(943):287-291 [FREE Full text] [Medline: 15138320]
45. Centis E, Marzocchi R, Di DS, Ciaravella MF, Marchesini G. The effect of lifestyle changes in non-alcoholic fatty liver disease. Dig Dis 2010;28(1):267-273. [doi: 10.1159/000282101] [Medline: 20460922]
46. Perrin A, Duggan M. Americans internet access - As internet use nears saturation for some groups, a look at patterns of adoption. 2000. Pew Research Center URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2015/06/2015-06-26_internet-usage-across-demographics-discover_FINAL.pdf [accessed 2018-06-11] [WebCite Cache ID 705lOu11M]
47. Fox, S. 2005. Health Information Online URL: http://www.pewinternet.org/files/old-media/Files/Reports/2005/PIP_Healthtopics_May05.pdf.pdf [accessed 2018-06-11] [WebCite Cache ID 705lYT6wS]
48. 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]
49. Williams DM, Matthews CE, Rutt C, Napolitano MA, Marcus BH. Interventions to increase walking behavior. Med Sci Sports Exerc 2008 Jul;40(7 Suppl):S567-S573 [FREE Full text] [doi: 10.1249/MSS.0b013e31817c7006] [Medline: 18562974]
50. Frensham LJ, Zarnowiecki DM, Parfitt G, King S, Dollman J. The experiences of participants in an innovative online resource designed to increase regular walking among rural cancer survivors: a qualitative pilot feasibility study. Support Care Cancer 2014 Jul;22(7):1923-1929. [doi: 10.1007/s00520-014-2177-4] [Medline: 24573604]
51. Lee MK, Yun YH, Park H, Lee ES, Jung KH, Noh D. A Web-based self-management exercise and diet intervention for breast cancer survivors: pilot randomized controlled trial. Int J Nurs Stud 2014 Dec;51(12):1557-1567. [doi: 10.1016/j.ijnurstu.2014.04.012] [Medline: 24856854]
52. Pfirrmann D, Tug S, Brosteanu O, Mehdorn M, Busse M, Grimminger PP, et al. Internet-based perioperative exercise program in patients with Barrett's carcinoma scheduled for esophagectomy [iPEP - study] a prospective randomized-controlled trial. BMC Cancer 2017 Jun 14;17(1):413 [FREE Full text] [doi: 10.1186/s12885-017-3400-8] [Medline: 28615010]
53. Pfirrmann D, Haller N, Huber Y, Jung P, Lieb K, Gockel I, et al. Applicability of a Web-Based, Individualized Exercise Intervention in Patients With Liver Disease, Cystic Fibrosis, Esophageal Cancer, and Psychiatric Disorders: Process Evaluation of 4 Ongoing Clinical Trials. JMIR Res Protoc 2018 May 22;7(5):e106 [FREE Full text] [doi: 10.2196/resprot.8607] [Medline: 29789277]
54. ACSM. In: Pescatello LS, Arena R, Riebe D, Thompson PD, editors. ACSM's Guidelines for Exercise Testing and Prescription. Philadelphia, Pennsylvania: Lippincott Williams & Wilkins; 2013.
55. 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]
56. Borg G. Perceived exertion as an indicator of somatic stress. Scand J Rehabil Med 1970;2(2):92-98. [Medline: 5523831]
57. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
58. World Health Organization. Global Recommendations on Physical Activity for Health. World Health Organization 2010:1-60. [Medline: 26180873]
59. Johnson NA, George J. Fitness versus fatness: moving beyond weight loss in nonalcoholic fatty liver disease. Hepatology 2010 Jul;52(1):370-381. [doi: 10.1002/hep.23711] [Medline: 20578153]
60. Cortez-Pinto H, Machado M. Impact of body weight, diet and lifestyle on nonalcoholic fatty liver disease. Expert Rev Gastroenterol Hepatol 2008 Apr;2(2):217-231. [doi: 10.1586/17474124.2.2.217] [Medline: 19072357]
61. Andersen T, Gluud C, Franzmann MB, Christoffersen P. Hepatic effects of dietary weight loss in morbidly obese subjects. J Hepatol 1991 Mar;12(2):224-229. [Medline: 2051001]

62. Pugh CJA, Spring VS, Kemp GJ, Richardson P, Shojaee-Moradie F, Umpleby AM, et al. Exercise training reverses endothelial dysfunction in nonalcoholic fatty liver disease. Am J Physiol Heart Circ Physiol 2014 Nov 01;307(9):H1298-H1306 [FREE Full text] [doi: 10.1152/ajpheart.00306.2014] [Medline: 25193471]
63. Hallsworth K, Fattakhova G, Hollingsworth KG, Thoma C, Moore S, Taylor R, et al. Resistance exercise reduces liver fat and its mediators in non-alcoholic fatty liver disease independent of weight loss. Gut 2011 Sep;60(9):1278-1283 [FREE Full text] [doi: 10.1136/gut.2011.242073] [Medline: 21708823]
64. Keating SE, Hackett DA, Parker HM, O'Connor HT, Gerofi JA, Sainsbury A, et al. Effect of aerobic exercise training dose on liver fat and visceral adiposity. J Hepatol 2015 Jul;63(1):174-182. [doi: 10.1016/j.jhep.2015.02.022] [Medline: 25863524]
65. Levine JA, Eberhardt NL, Jensen MD. Role of nonexercise activity thermogenesis in resistance to fat gain in humans. Science 1999 Jan 08;283(5399):212-214 [FREE Full text] [Medline: 9880251]
66. Blair SN. Evidence for success of exercise in weight loss and control. Ann Intern Med 1993 Oct 01;119(7 Pt 2):702-706. [Medline: 8363201]
67. Takahashi A, Abe K, Usami K, Imaizumi H, Hayashi M, Okai K, et al. Simple Resistance Exercise helps Patients with Non-alcoholic Fatty Liver Disease. Int J Sports Med 2015 Oct;36(10):848-852. [doi: 10.1055/s-0035-1549853] [Medline: 26090879]
68. Cuthbertson DJ, Shojaee-Moradie F, Sprung VS, Jones H, Pugh CJA, Richardson P, et al. Dissociation between exercise-induced reduction in liver fat and changes in hepatic and peripheral glucose homoeostasis in obese patients with non-alcoholic fatty liver disease. Clin Sci (Lond) 2016 Jan;130(2):93-104. [doi: 10.1042/CS20150447] [Medline: 26424731]
69. Johnson NA, Sachinwalla T, Walton DW, Smith K, Armstrong A, Thompson MW, et al. Aerobic exercise training reduces hepatic and visceral lipids in obese individuals without weight loss. Hepatology 2009 Oct;50(4):1105-1112. [doi: 10.1002/hep.23129] [Medline: 19637289]
70. Sullivan S, Kirk EP, Mittendorfer B, Patterson BW, Klein S. Randomized trial of exercise effect on intrahepatic triglyceride content and lipid kinetics in nonalcoholic fatty liver disease. Hepatology 2012 Jun;55(6):1738-1745 [FREE Full text] [doi: 10.1002/hep.25548] [Medline: 22213436]
71. Dunbar CC, Robertson RJ, Baun R, Blandin MF, Metz K, Burdett R, et al. The validity of regulating exercise intensity by ratings of perceived exertion. Med Sci Sports Exerc 1992 Jan;24(1):94-99. [Medline: 1549002]
72. Scherr J, Wolfarth B, Christle JW, Pressler A, Wagenpfeil S, Halle M. Associations between Borg's rating of perceived exertion and physiological measures of exercise intensity. Eur J Appl Physiol 2013 Jan;113(1):147-155. [doi: 10.1007/s00421-012-2421-x] [Medline: 22615009]
73. Shamsoddini A, Sobhani V, Ghamar CME, Alavian SM, Zaree A. Effect of Aerobic and Resistance Exercise Training on Liver Enzymes and Hepatic Fat in Iranian Men With Nonalcoholic Fatty Liver Disease. Hepat Mon 2015 Oct;15(10):e31434 [FREE Full text] [doi: 10.5812/hepatmon.31434] [Medline: 26587039]
74. Perri MG, Anton SD, Durning PE, Ketterson TU, Sydeman SJ, Berlant NE, et al. Adherence to exercise prescriptions: effects of prescribing moderate versus higher levels of intensity and frequency. Health Psychol 2002 Sep;21(5):452-458. [Medline: 12211512]
75. Golabi P, Locklear CT, Austin P, Afdhal S, Byrns M, Gerber L, et al. Effectiveness of exercise in hepatic fat mobilization in non-alcoholic fatty liver disease: Systematic review. World J Gastroenterol 2016 Jul 21;22(27):6318-6327 [FREE Full text] [doi: 10.3748/wjg.v22.i27.6318] [Medline: 27468220]
76. St George A, Bauman A, Johnston A, Farrell G, Chey T, George J. Independent effects of physical activity in patients with nonalcoholic fatty liver disease. Hepatology 2009 Jul;50(1):68-76. [doi: 10.1002/hep.22940] [Medline: 19444870]
77. Zelber-Sagi S, Nitzan-Kaluski D, Goldsmith R, Webb M, Zvibel I, Goldiner I, et al. Role of leisure-time physical activity in nonalcoholic fatty liver disease: a population-based study. Hepatology 2008 Dec;48(6):1791-1798. [doi: 10.1002/hep.22525] [Medline: 18972405]
78. Hsieh SD, Yoshinaga H, Muto T, Sakurai Y. Regular physical activity and coronary risk factors in Japanese men. Circulation 1998 Feb 24;97(7):661-665 [FREE Full text] [Medline: 9495301]
79. Perseghin G, Lattuada G, De Cobelli F, Ragogna F, Ntali G, Esposito A, et al. Habitual physical activity is associated with intrahepatic fat content in humans. Diabetes Care 2007 Mar;30(3):683-688. [doi: 10.2337/dc06-2032] [Medline: 17327341]
80. Trenell MI. Sedentary behaviour, physical activity, and NAFLD: Curse of the chair. J Hepatol 2015 Nov;63(5):1064-1065. [doi: 10.1016/j.jhep.2015.08.009] [Medline: 26297919]
81. Vandelanotte C, Sugiyama T, Gardiner P, Owen N. Associations of leisure-time internet and computer use with overweight and obesity, physical activity and sedentary behaviors: cross-sectional study. J Med Internet Res 2009 Jul 27;11(3):e28 [FREE Full text] [doi: 10.2196/jmir.1084] [Medline: 19666455]
82. Montesi L, Moscatiello S, Malavolti M, Marzocchi R, Marchesini G. Physical activity for the prevention and treatment of metabolic disorders. Intern Emerg Med 2013 Dec;8(8):655-666. [doi: 10.1007/s11739-013-0953-7] [Medline: 23657989]
83. Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. J Med Internet Res 2009;11(2):e15 [FREE Full text] [doi: 10.2196/jmir.1151] [Medline: 19632969]

84. Ritterband LM, Gonder-Frederick LA, Cox DJ, Clifton AD, West RW, Borowitz SM. Internet interventions: In review, in use, and into the future. Professional Psychology: Research and Practice 2003;34(5):527-534. [doi: 10.1037/0735-7028.34.5.527]
85. Casey D, De Civita M, Dasgupta K. Understanding physical activity facilitators and barriers during and following a supervised exercise programme in Type 2 diabetes: a qualitative study. Diabet Med 2010 Jan;27(1):79-84. [doi: 10.1111/j.1464-5491.2009.02873.x] [Medline: 20121893]
86. 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] [Medline: 21694556]
87. Johnson NA, Keating SE, George J. Exercise and the liver: implications for therapy in fatty liver disorders. Semin Liver Dis 2012 Feb;32(1):65-79. [doi: 10.1055/s-0032-1306427] [Medline: 22418889]
88. Levine JA. Sick of sitting. Diabetologia 2015 Aug;58(8):1751-1758. [doi: 10.1007/s00125-015-3624-6] [Medline: 26003325]
89. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
90. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
91. Marcus BH, Lewis BA, Williams DM, Dunsiger S, Jakicic JM, Whiteley JA, et al. A comparison of Internet and print-based physical activity interventions. Arch Intern Med 2007 May 14;167(9):944-949. [doi: 10.1001/archinte.167.9.944] [Medline: 17502536]
92. Afdhal NH. Management of nonalcoholic fatty liver disease: a 60-year-old man with probable nonalcoholic fatty liver disease: weight reduction, liver biopsy, or both? JAMA 2012 Aug 08;308(6):608-616. [doi: 10.1001/jama.2012.8402] [Medline: 22760235]
93. Bantum EO, Albright CL, White KK, Berenberg JL, Layi G, Ritter PL, et al. Surviving and thriving with cancer using a Web-based health behavior change intervention: randomized controlled trial. J Med Internet Res 2014;16(2):e54 [FREE Full text] [doi: 10.2196/jmir.3020] [Medline: 24566820]
94. Hatchett A, Hallam JS, Ford MA. Evaluation of a social cognitive theory-based email intervention designed to influence the physical activity of survivors of breast cancer. Psychooncology 2013 Apr;22(4):829-836. [doi: 10.1002/pon.3082] [Medline: 22573338]

Abbreviations

BMI: body mass index
HELP: Hepatic Inflammation and Physical Performance in Patients With NASH
NAFLD: nonalcoholic fatty liver disease
NASH: nonalcoholic steatohepatitis
VO_{2max}: maximum volume of oxygen
VO_{2peak}: peak volume of oxygen

Edited by G Eysenbach; submitted 13.06.18; peer-reviewed by C Carrion, L Wesselman; comments to author 06.09.18; revised version received 25.09.18; accepted 04.10.18; published 02.01.19.

Please cite as:
Pfirrmann D, Huber Y, Schattenberg JM, Simon P
Web-Based Exercise as an Effective Complementary Treatment for Patients With Nonalcoholic Fatty Liver Disease: Intervention Study
J Med Internet Res 2019;21(1):e11250
URL: http://www.jmir.org/2019/1/e11250/
doi:10.2196/11250
PMID:30602434

Review

Web-Based Digital Health Interventions for Weight Loss and Lifestyle Habit Changes in Overweight and Obese Adults: Systematic Review and Meta-Analysis

Alline M Beleigoli[1,2,3], MD, PhD; Andre Q Andrade[4], MD, PhD; Alexandre G Cançado[1], MD; Matheus NL Paulo[1], MD; Maria De Fátima H Diniz[1], MD, PhD; Antonio L Ribeiro[1,5], MD, PhD

[1]Internal Medicine Department, Faculty of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[2]Department of Medicine, University of Adelaide, Adelaide, Australia

[3]Flinders Digital Health Research Centre, College of Nursing and Health Sciences, Flinders University, Adelaide, Australia

[4]Quality Use of Medicines and Pharmacy Research Center, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, Australia

[5]Telehealth Center, Hospital das Clinicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Corresponding Author:
Alline M Beleigoli, MD, PhD
Internal Medicine Department
Faculty of Medicine
Universidade Federal de Minas Gerais
Avenida Professor Alfredo Balena
190 Santa Efigenia
Belo Horizonte, 30130100
Brazil
Phone: 55 3134099746
Email: abeleigoli@gmail.com

Abstract

Background: Obesity is a highly prevalent condition with important health implications. Face-to-face interventions to treat obesity demand a large number of human resources and time, generating a great burden to individuals and health system. In this context, the internet is an attractive tool for delivering weight loss programs due to anonymity, 24-hour-accessibility, scalability, and reachability associated with Web-based programs.

Objective: We aimed to investigate the effectiveness of Web-based digital health interventions, excluding hybrid interventions and non-Web-based technologies such as text messaging, short message service, in comparison to nontechnology active or inactive (wait list) interventions on weight loss and lifestyle habit changes in individuals with overweight and obesity.

Methods: We searched PubMed or Medline, SciELO, Lilacs, PsychNet, and Web of Science up to July 2018, as well as references of previous reviews for randomized trials that compared Web-based digital health interventions to offline interventions. Anthropometric changes such as weight, body mass index (BMI), waist, and body fat and lifestyle habit changes in adults with overweight and obesity were the outcomes of interest. Random effects meta-analysis and meta-regression were performed for mean differences (MDs) in weight. We rated the risk of bias for each study and the quality of evidence across studies using the Grades of Recommendation, Assessment, Development, and Evaluation approach.

Results: Among the 4071 articles retrieved, 11 were included. Weight (MD −0.77 kg, 95% CI −2.16 to 0.62; 1497 participants; moderate certainty evidence) and BMI (MD −0.12 kg/m2; 95% CI −0.64 to 0.41; 1244 participants; moderate certainty evidence) changes were not different between Web-based and offline interventions. Compared to offline interventions, digital interventions led to a greater short-term (<6 months follow-up) weight loss (MD −2.13 kg, 95% CI −2.71 to −1.55; 393 participants; high certainty evidence), but not in the long-term (MD −0.17 kg, 95% CI −2.10 to 1.76; 1104 participants; moderate certainty evidence). Meta-analysis was not possible for lifestyle habit changes. High risk of attrition bias was identified in 5 studies. For weight and BMI outcomes, the certainty of evidence was moderate mainly due to high heterogeneity, which was mainly attributable to control group differences across studies (R^2=79%).

Conclusions: Web-based digital interventions led to greater short-term but not long-term weight loss than offline interventions in overweight and obese adults. Heterogeneity was high across studies, and high attrition rates suggested that engagement is a major issue in Web-based interventions.

(J Med Internet Res 2019;21(1):e298) doi:10.2196/jmir.9609

KEYWORDS

internet; mobile phone; meta-analysis; obesity; telemedicine

Introduction

Facing the global obesity epidemic is a major public health challenge [1]. The prevalence of obesity has nearly doubled over the last 30 years [1]. Obesity is associated with an increased risk for type 2 diabetes, hypertension, dyslipidemia, cardiovascular diseases, musculoskeletal disorders, psychological stress, and certain types of cancer. All these morbidities significantly increase mortality and reduce quality of life [2].

Obesity treatment involves a systemic approach with both individual and environmental strategies [3]. The individual interventions are usually delivered face-to-face, which generate high demands for individuals, due to their prolonged course, and a great burden to the health care system due to the high prevalence of obesity [4]. Despite such efforts, the effectiveness of obesity interventions on weight loss is only modest, particularly in the long-term [5].

In this context, Web-based digital technology can be a particularly interesting tool for the treatment of overweight and obesity due to its capacity for reaching a large number of people even in remote areas on a 24-hour per 7-day regimen. Delivering weight loss interventions on the Web allows targeting a larger number of people compared to face-to-face interventions and might be less time consuming and more cost-effective for professionals and patients [6]. Previous reviews have shown a modest superiority of digital interventions in comparison to offline interventions with regards to weight loss [7,8]. However, as these reviews included studies that investigated hybrid interventions both in the intervention (eg, Web-based plus short message service text messages) and control groups (ie, face-to-face plus technology-based interventions), the effect of interventions that use only Web-based delivery is not known.

Our aim was to conduct a systematic review and meta-analysis of randomized controlled trials to investigate the effect of Web-based digital interventions in comparison to real-world interventions on anthropometric measures and changes in dietary and physical activity habits in individuals with overweight and obesity.

Methods

Systematic Review

For the purpose of this review, PubMed or Medline, SciELO, Lilacs, PsychNet, and Web of Science electronic databases were searched up to July 1, 2018. No language restrictions were applied. We searched both for indexed terms and terms in titles or abstracts that corresponded to the following search pattern in PubMed or Medline: (overweight OR obes*) AND (web OR technology OR internet OR computers OR “social media” OR online).

Studies were eligible if they reported data on randomized controlled trials, which recruited adults (≥18 years) with overweight and obesity (body mass index [BMI] ≥25 kg/m^2) into a Web-based digital intervention (accessed by browser or Web-based application, regardless of device) versus offline or in-person (face-to-face) interventions. Studies that did not apply any active interventions (wait list) in the control group were also included. Exclusion criteria comprised studies in which overweight and obesity were not a primary selection criterion or those in which the predefined outcomes were not reported. Additionally, studies that included children, adolescents, or pregnant women were excluded. Trials of hybrid interventions (Web-based digital interventions plus face-to-face interventions or other technology-based interventions, such as mobile short message service text messages or digital interventions plus offline interventions) and those that included digital interventions in the control group were also excluded. Moreover, studies evaluating the prevention of weight regain after a previous intervention and those that did not report the predefined outcomes of interest were not included. Multiple reports from the same study were considered as a single one. We considered changes in anthropometric measures and in dietary and physical activity habits as the outcomes of interest.

Two reviewers (AGC and MNLP) independently carried out the selection of the studies according to the predefined eligibility criteria. Any disagreement between them was evaluated by 2 other authors (AMB and AQA). AMB and AQA independently extracted data from reports based on a predefined data extraction form. Any disagreement between them was evaluated by either MdFHD or ALR. When some information was not clear in the report, authors were contacted by email. Hand search was performed in the references of previously published reviews.

Quality of the Evidence

We used the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) methodology [9] to assess the quality of evidence retrieved by the systematic review. This consists of evaluating the risk of selection (randomization and allocation), performance (blinding of participants and personnel), detection (blinding of outcome assessment), attrition (incomplete outcome data), and reporting (selective reporting) bias of individual studies. In addition, the GRADE methodology suggests assessment of indirectness, inconsistency, imprecision and publication bias of the evidence overall in order to grade the level of the evidence retrieved.

Meta-Analysis

We used a random effects model to calculate summary mean differences (MDs) and 95% CIs for 1 unit change in weight (kg), BMI (kg/m^2), waist (cm), or body fat (%). In cases where both "per protocol" and intention-to-treat results were provided, the latter were used to calculate summary MDs. For dietary and physical activity habits, we found a great diversity in the instruments used to measure changes among the groups. This finding precluded meta-analysis, and we performed only qualitative analysis of these outcomes. We used the Cochrane Review Manager software for these analyses [10].

A random effects meta-regression model was used to determine whether the type of control group (with and without active intervention) was a source of heterogeneity among studies. We performed sensitivity analyses according to the length of follow-up (<6 and ≥6 months) and the type of control intervention (presence or not of a nondigital intervention in the control group). We used *Comprehensive Meta-Analysis* software version 3 for this analysis [11].

Results

Systematic Review

The search strategy resulted in 4071 articles. After exclusions, as shown in Figure 1, a total of 11 studies [12-22] that analyzed data from 1525 participants were retrieved. Female sex was predominant in most of the studies. The age of the participants varied from 18 to 65 years. Most of the studies excluded participants with comorbidities and pregnancy as well as participants who were engaged in other weight loss programs. Unhealthy lifestyle habits were not an inclusion criterion in any of the retrieved studies. Recruitment settings varied among the studies and included community populations, physician-referred patients, company employees, and university students or staff. Other characteristics of the studies retrieved are depicted in Multimedia Appendix 1.

Interventions were predominantly delivered via internet browsers, except 2 that were delivered by a smartphone app [12,16]. Diverse behavioral strategies, such as goal setting, self-monitoring and management, social support, modeling, and feedback were applied in the studies. The control groups received either no intervention (wait list) or usual face-to-face interventions (Table 1).

Quality of the Evidence

With regard to the risk of bias of individual studies (Figures 2 and 3), all of the studies reported a sequence generation randomization process. Allocation was not concealed in 3 studies [12,18,20]. As expected in this type of intervention, blinding of participants and personnel was not feasible, whereas blinding of the assessor was not reported in 5 [15,17,19-22]. Moreover, high follow-up attrition rates were a common finding. Moreover, 7 of the 11 retrieved studies showed ≥20% losses to follow-up, and unbalanced losses (intervention>control group) were present in 5 of the 11 studies [12,14,16,18,19].

The quality of the evidence retrieved by the GRADE methodology was considered moderate for the primary outcomes of this review (weight and BMI change), as shown in the summary of findings table (Multimedia Appendix 2). Although indirectness, imprecision, and publication bias were not major issues in this body of evidence, heterogeneity (I^2=94%; P<.001, for weight loss as the outcome) was high and explained mainly (R^2=0.79) by differences in the type of control group as shown by meta-regression analysis.

Anthropometric Measures

Absolute weight and BMI changes were reported in 10 [12-15,17-22] and 9 [12-20] studies, respectively. Overall, changes in weight (MD −0.77 kg; 95% CI −2.16 to 0.62; Figure 2) and MDs in BMI (MD −0.12 kg/m^2; 95% CI −0.64 to 0.41; Figure 3) were not significantly different between the digital interventions and the offline interventions. Only 2 studies reported results on waist circumference [12,14]. There was no difference between the intervention and control groups for this outcome (−0.54 cm; 95% CI −5.17 to 4.10), as shown in Figure 4. Only 1 study reported changes in percent body fat and did not find a significant difference between the intervention and control groups (−1.40%; 95% CI −2.93 to 0.13) [13].

Lifestyle Habits and Other Outcomes

Among the 11 studies, 8 reported outcomes on dietary or physical activity habit changes [12,14-17,19-21]. However, the instruments used to measure qualitative and quantitative dietary and physical activity characteristics were very different across the studies. This precluded us to perform a quantitative review of these outcomes. Most of the studies reported that there was no significant difference between the intervention and control groups, except for dietary habits in 3 of the studies (Table 2) [14,16,21]. Moreover, 5 studies [14,16,17,20,22] reported data on substitutive measures of cardiovascular morbidity—blood pressure, glucose metabolism, or cholesterol. None of them found any difference between the intervention and control groups. None of the 11 studies investigated hard endpoints, such as cardiovascular morbidity and mortality.

There was no difference between the groups in terms of quality of life in the 3 studies that assessed it [18-20]. The Web-based intervention was cost-effective in comparison to a 6-month in-person intervention in 1 [18] of the 3 studies that evaluated cost-effectiveness [18-20].

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analysis flowchart.

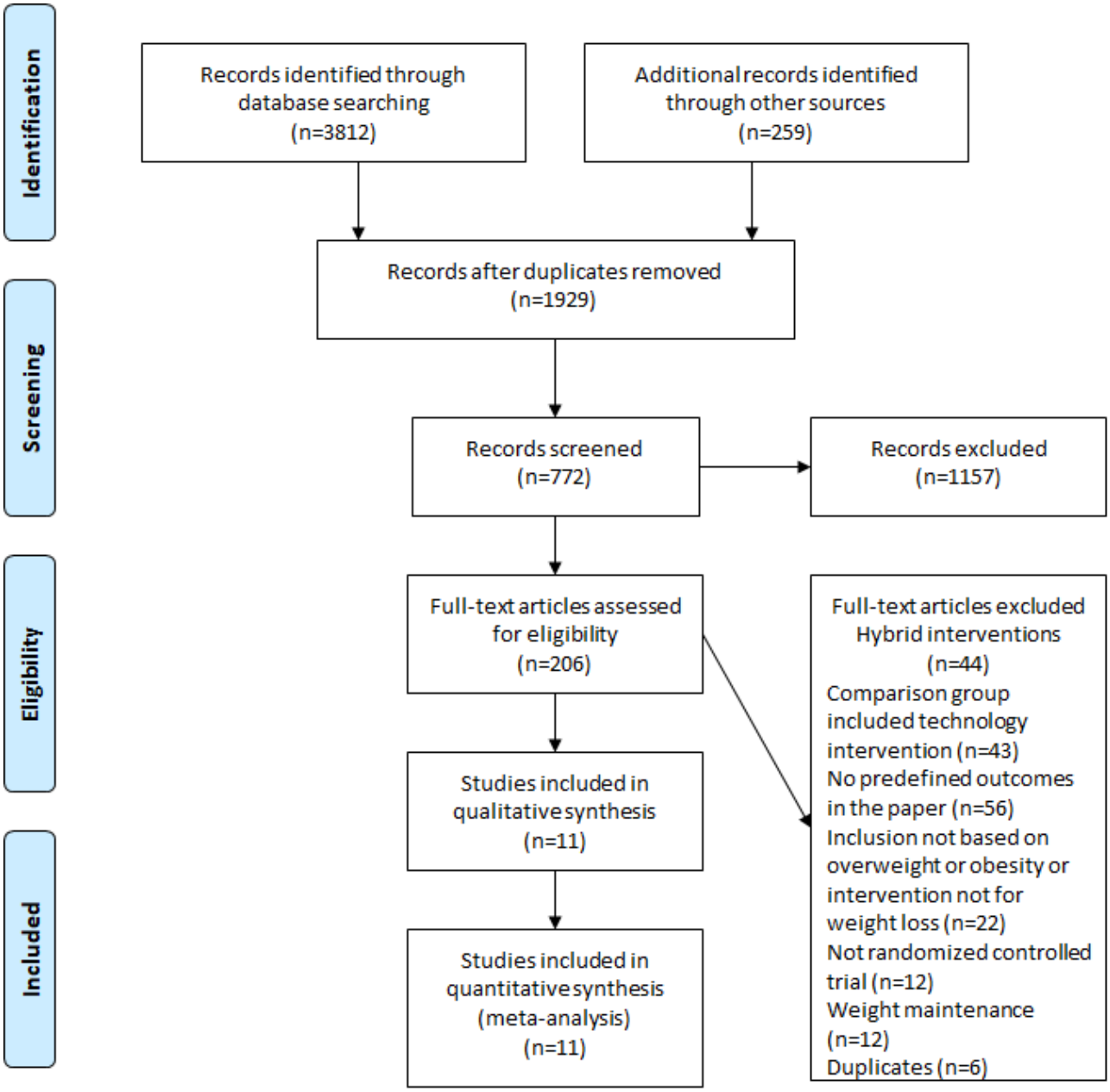
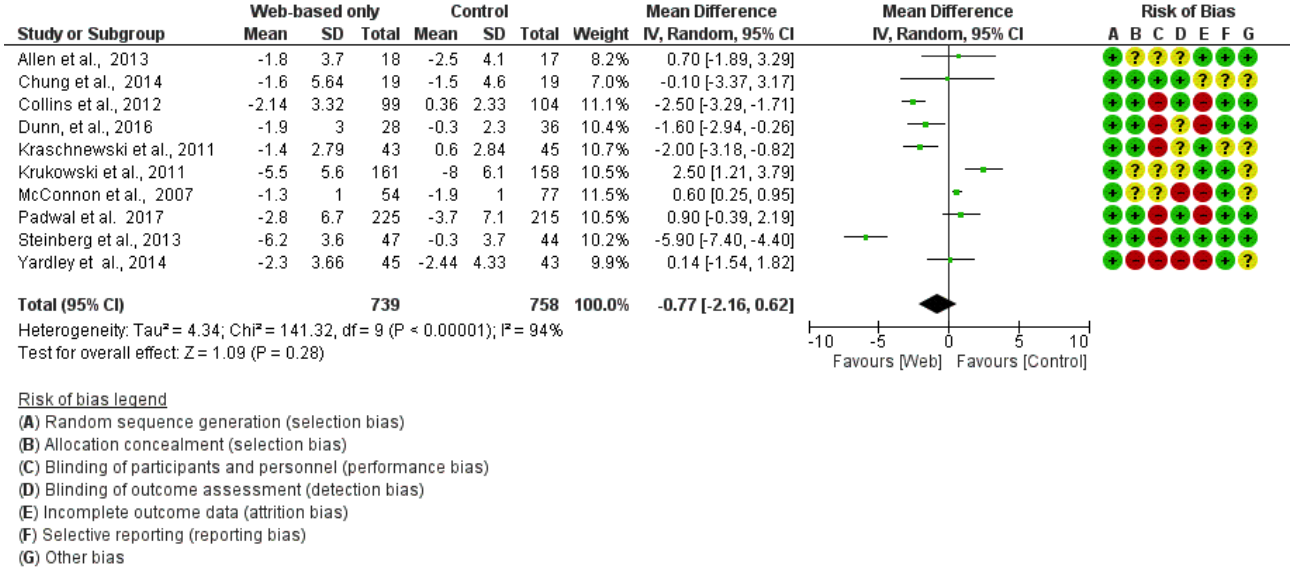


Table 1. Details of the intervention and control groups.

Study	Behavioral theory	Behavioral strategy	Technology strategy	Type of control intervention	Behavioral strategy in control group	Tools used in control group
Allen et al 2013 [12]	Social cognitive theory, behavioral self-management, and motivational interviewing counseling	Self-management, mindful empowerment, and feedback	Web-based smartphone app	In person	Coaching and goal setting	N/A[a]
Chung et al 2014 [13]	Not specified	Self-monitoring, knowledge, personalized feedback	Image-based electronic portal and caloric calculator	Paper food diary	Self-monitoring	Paper diary
Collins et al 2012 [14]	Social cognitive theory	Self-efficacy, goal setting, and self-monitoring of weight, body measurements, exercise, and diet; outcome expectations (knowledge-based web components); modeling; and social support	Website and telephone contact	Wait list	N/A	N/A
Dunn et al 2016 [15]	Theory of planned behavior, mindfulness, and small steps to change	Web-based lessons of the Eat Smart, Move More, Weigh Less, an evidence-based, 15-week, adult weight management program	Website	Wait list	Theory of planned behavior, mindfulness, and small steps to change	Eat Smart, Move More, Weigh Less, an evidence-based, 15-week, adult weight management program
Hurkmans et al 2018 [16]	Not specified	Knowledge, self-monitoring, help button	Mobile app	In person	Self-monitoring, action planning, relapse prevention	N/A
Kraschnewski 2011 [17]	Positive deviance framework	Modeling, goal setting, knowledge, personalized feedback, self-monitoring	Videos	Wait list	N/A	N/A
Krukowski et al 2011 [18]	Not specified	Knowledge, self-monitoring, stimulus control, problem solving, goal setting, relapse prevention, and assertiveness training	Chat group, pedometers, and website platform	In person	Knowledge, self-monitoring, stimulus control, problem solving, goal setting, relapse prevention, and assertiveness training	Group sessions and printed information on dietary intake and physical activity
McConnon et al 2007 [19]	Not specified	Counseling, personal feedback	Not specified	Primary-care based printed information	Not available	Printed information
Padwal et al 2017 [20]	Not specified	Knowledge, self-management, self-monitoring, goal setting, stress management	Website	In person	Knowledge, self-management, self-monitoring, goal setting, and stress management	N/A
Steinberg et al 2013 [21]	Not specified	Self-monitoring, knowledge and skills (portion control, restaurant eating, structured exercise, problem solving, stimulus control, and relapse prevention)	Cellular-connected "smart" scale for daily weighing, website, and email	Wait list	N/A	N/A
Yardley et al 2014 [22]	Cognitive behavioral theory	Skills, self-regulation, and feedback	Website lessons, challenges, and email	In person	Not available	N/A

[a]N/A: not applicable.

Figure 2. Meta-analysis results for mean weight change (kg) in Web-based-only versus offline interventions. df: degrees of freedom; IV: interval variable; random: random effects model.



Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Blinding of participants and personnel (performance bias)
(D) Blinding of outcome assessment (detection bias)
(E) Incomplete outcome data (attrition bias)
(F) Selective reporting (reporting bias)
(G) Other bias

Figure 3. Meta-analysis results for mean body mass index change (kg/m2) in Web-based-only versus offline interventions. df: degrees of freedom; IV: interval variable; random: random effects model.

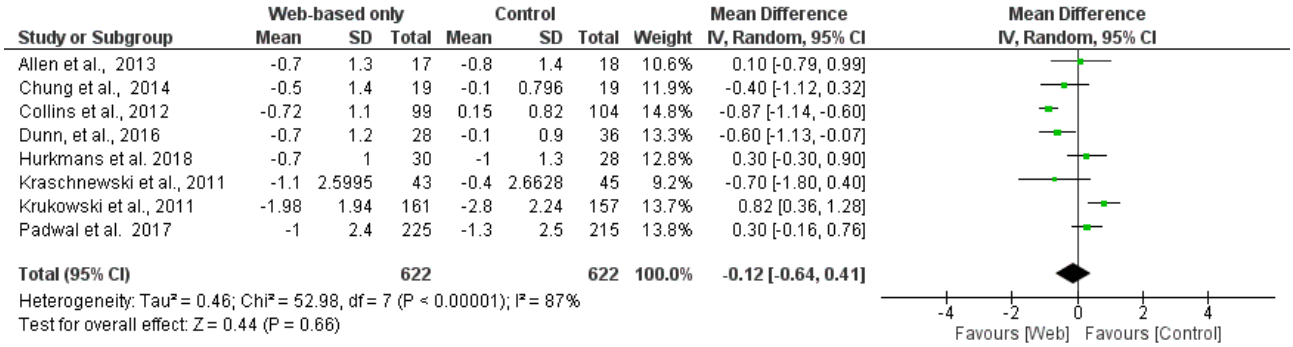


Figure 4. Meta-analysis results for mean waist change (cm) in Web-based-only versus offline interventions for studies with a specific intervention in the control group. df: degrees of freedom; IV: interval variable; random: random effects model.

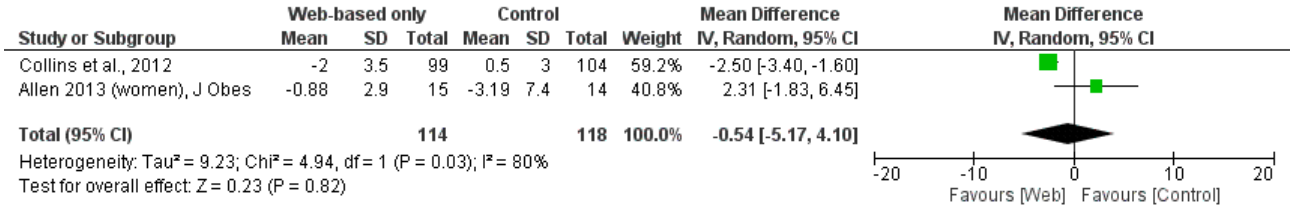


Table 2. Differences of lifestyle habits between the intervention and control groups.

Study	Dietary caloric intake	Dietary quality	Physical activity
Allen et al 2013 [12]	No difference	N/A[a]	No difference
Chung et al [13]	N/A	N/A	N/A
Collins et al 2012 [14]	Lower in the intervention group	No difference	No difference
Dunn et al 2016 [15]	No difference	No difference	No difference
Hurkmans et al 2018 [16]	Lower in control group	No difference	No difference
Kraschnewski et al, 2011 [17]	No difference	No difference	No difference
Krukowski et al [18]	N/A	No difference	No difference
McConnon et al [19]	No difference	No difference	No difference
Padwal et al 2017 [20]	N/A	N/A	N/A
Steinberg et al 2013 [21]	Lower in the intervention group	No difference	No difference
Yardley et al 2014 [22]	N/A	N/A	N/A

[a]N/A: not available.

Sensitivity Analyses

In the subgroup of studies in which there was an active intervention in the control group, there was a significant difference between Web-based interventions and nontechnology interventions regarding weight loss (MD 0.82 kg; 95% CI 0.06 to 1.59; Figure 5). When the analysis was restricted to the subgroup of studies that did not have any intervention in the control group, the Web-based intervention was superior to control (MD −2.14 kg; 95% CI −2.65 to −1.64; Figure 6).

When studies were analyzed according to the length of follow-up, there was greater weight loss (MD −2.13 kg; 95% CI −2.71 to −1.55) in the Web-based intervention group than in the offline intervention group in the subgroup of studies with <6 months of follow-up, whereas there was no difference between the intervention and control groups in the subgroup of studies with ≥6 months of follow-up (MD −0.17 kg; 95% CI −2.10 to 1.76), as shown in Figures 7 and 8, respectively.

Figure 5. Meta-analysis results for mean weight change (kg) in Web-based-only versus active nontechnology interventions in the control group. df: degrees of freedom; IV: interval variable; random: random effects model.

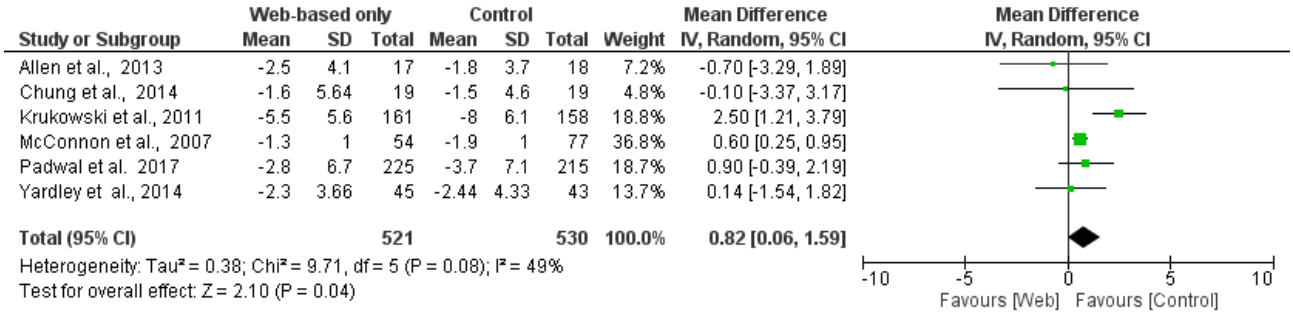


Figure 6. Meta-analysis results for mean weight change (kg) in Web-based-only versus nonactive interventions (wait list) in the control group. df: degrees of freedom; IV: interval variable; random: random effects model.

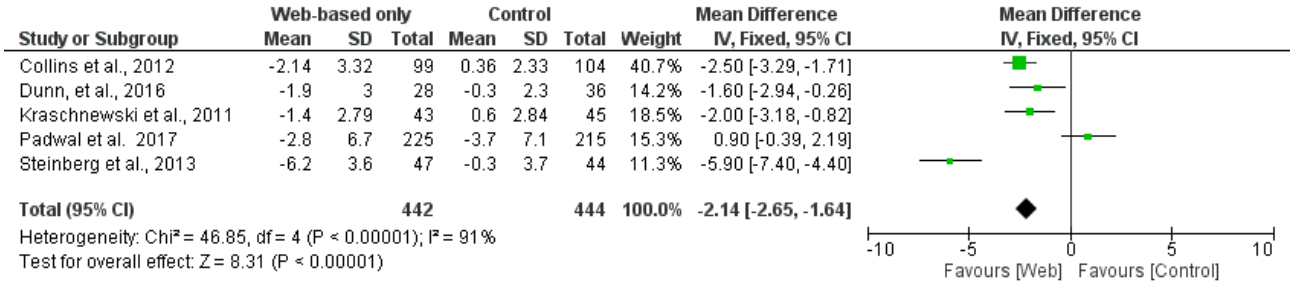


Figure 7. Meta-analysis results for mean weight change (kg) in Web-based-only versus offline interventions for studies with <6 months follow-up duration. df: degrees of freedom; IV: interval variable; random: random effects model.

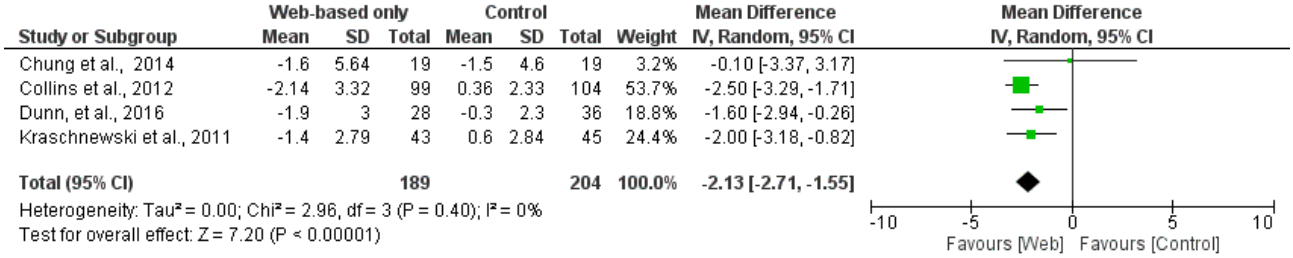
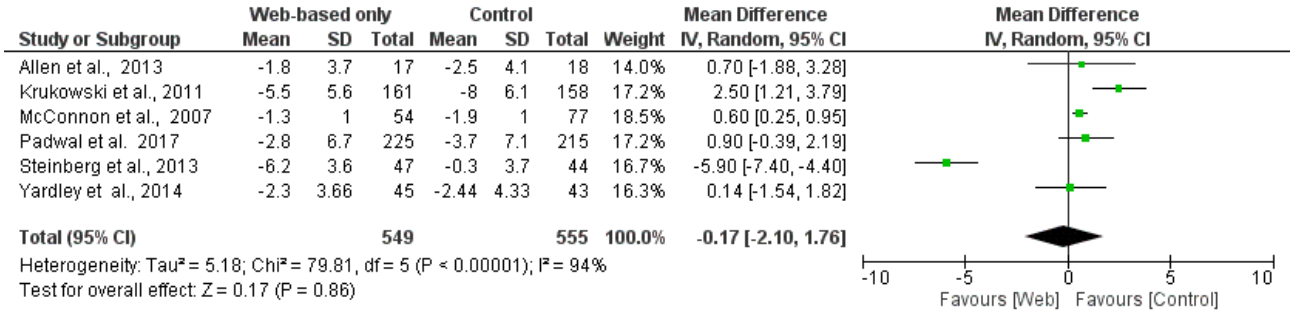


Figure 8. Meta-analysis results for mean weight change (kg) in Web-based-only versus offline interventions for studies with ≥6 months follow-up duration. df: degrees of freedom; IV: interval variable; random: random effects model.



Discussion

Principal Findings

In this meta-analysis, we found that the use of Web-based digital interventions exclusively was not superior to the use of offline interventions in terms of weight or BMI loss in individuals with overweight and obesity except in the short-term. These findings were based on moderate-quality evidence. Changes in dietary and physical habits of overweight and obese individuals were not different between these 2 types of intervention either.

The findings of superiority of the intervention in comparison to the control for short-term but not long-term weight loss suggest that long-term use and adherence to digital interventions are important issues to consider when planning this kind of intervention. Moreover, the superiority of digital intervention in the subgroup of studies that had no specific intervention in the comparison group suggests that this tool might be more valuable to induce weight loss in patients who do not have access to any kind of in-person intervention.

Intervention-induced weight loss was of small clinical significance. This happened even in studies with a short-term follow-up. Low engagement to the interventions delivered by the Web-based tools might explain these modest results and might be a proxy for the low motivation of participants [23]. These modest results also highlight the need to investigate the components and tools of Web-based platforms that lead to the maintenance of users' motivation, interest, and participation, which play a key role in enhancing adherence to healthy behaviors.

A great diversity of behavioral techniques was found in the intervention groups across studies. Behavioral strategies with multiple components comprised most of the intervention strategies. This makes it difficult to infer which components are more effective in promoting weight loss and change of health habits and precluded us from identifying whether the results were due to differences in the nature of the interface (Web-based vs face-to-face) or in the behavioral strategy. Additionally, the principles of the interventions applied in the control group were not similar to those applied in the intervention group within each study.

High risk of attrition bias was identified in 5 of the 11 studies. Although most of them followed up participants in the short-term (less than 6 months), loss of ≥20% of participants over the follow-up period was common both in the intervention and control groups. This suggests that Web-based interventions probably do not overcome the low adherence to treatments, which is commonly reported in obesity studies. Another issue of concern regarding the quality of the studies was the scarcity of data on hazardous outcomes related to the weight loss. Since appetite disorders as well as muscle and bone mass reduction may be consequences of weight loss, it was desirable that the studies had included these issues in the results. Differences in the type of control group (with and without intervention) explained a major part of the high heterogeneity found in the meta-analysis.

The thorough revision, which included 5 databases with no language restriction, is a major strength of this study. On the other hand, the high heterogeneity and high risk of attrition bias make recommendations of using Web-based interventions for individuals with overweight and obesity based on their effectiveness on weight loss of moderate certainty.

Conclusion

There is moderated certainty in our findings that Web-based digital health interventions are more effective than nontechnology interventions in promoting short-term but not long-term weight loss. Moreover, Web-based interventions do

not seem superior to nontechnology ones in terms of changes in dietary and physical activities. The high dropout rates in the retrieved studies contributed to a lowered quality of evidence and suggest that designing interventions that maintain participants' engagement and motivation over time might be fundamental to the success of digital interventions.

Acknowledgments

This study was supported with grants from the National Institute of Science and Technology for Health Technology Assessment (IATS) – CNPq/Brazil. AMB was was funded by Programa de Pós Doutorado (Prêmio Capes de Teses em Medicina I-2013) from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). ALR was supported by CNPq (research fellowship number 310679/2016–8) and FAPEMIG (PPM-00428-17).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Individual study characteristics.

[PDF File (Adobe PDF File), 37KB - jmir_v21i1e298_app1.pdf]

Multimedia Appendix 2

Summary of findings table according to the GRADE methodology.

[PDF File (Adobe PDF File), 48KB - jmir_v21i1e298_app2.pdf]

References

1. World Health Organization Fact Sheets. 2017 URL: http://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2018-08-14] [WebCite Cache ID 71ejzyw6n]
2. GBD 2015 Obesity Collaborators, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. N Engl J Med 2017 Dec 06;377(1):13-27 [FREE Full text] [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]
3. Bray GA, Frühbeck G, Ryan DH, Wilding JPH. Management of obesity. Lancet 2016 May 07;387(10031):1947-1956. [doi: 10.1016/S0140-6736(16)00271-3] [Medline: 26868660]
4. 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] [Medline: 25369490]
5. Douketis JD, Macie C, Thabane L, Williamson DF. Systematic review of long-term weight loss studies in obese adults: clinical significance and applicability to clinical practice. Int J Obes (Lond) 2005 Oct;29(10):1153-1167. [doi: 10.1038/sj.ijo.0802982] [Medline: 15997250]
6. Dávalos ME, French MT, Burdick AE, Simmons SC. Economic evaluation of telemedicine: review of the literature and research guidelines for benefit-cost analysis. Telemed J E Health 2009 Dec;15(10):933-948. [doi: 10.1089/tmj.2009.0067] [Medline: 19954346]
7. Arem H, Irwin M. A review of web-based weight loss interventions in adults. Obes Rev 2011 May;12(5):e236-e243 [FREE Full text] [doi: 10.1111/j.1467-789X.2010.00787.x] [Medline: 20804523]
8. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
9. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]
10. Review Manager (RevMan) computer program, 5 ed. Copenhagen: The Nordic Cochrane Centre: The Cochrane Collaboration; 2014.
11. Borenstein M, Hedges L, Higgins J, Rothstein H. Comprehensive Meta-Analysis Computer program. 3 ed. Englewood, NJ: Biostat; 2006.
12. Allen JK, Stephens J, Dennison Himmelfarb CR, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013:151597 [FREE Full text] [doi: 10.1155/2013/151597] [Medline: 24392223]

13. Chung LMY, Law QPS, Fong SSM, Chung JWY. Teledietetics improves weight reduction by modifying eating behavior: a randomized controlled trial. Telemed J E Health 2014 Jan;20(1):55-62 [FREE Full text] [doi: 10.1089/tmj.2013.0104] [Medline: 24205807]
14. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ, et al. A 12-week commercial web-based weight-loss program for overweight and obese adults: randomized controlled trial comparing basic versus enhanced features. J Med Internet Res 2012 Apr 25;14(2):e57 [FREE Full text] [doi: 10.2196/jmir.1980] [Medline: 22555246]
15. Dunn C, Olabode-Dada O, Whetstone L, Thomas C, Aggarwal S, Nordby K, et al. Using synchronous distance education to deliver a weight loss intervention: A randomized trial. Obesity (Silver Spring) 2016 Jan;24(1):44-50 [FREE Full text] [doi: 10.1002/oby.21315] [Medline: 26637964]
16. Hurkmans E, Matthys C, Bogaerts A, Scheys L, Devloo K, Seghers J. Face-to-Face Versus Mobile Versus Blended Weight Loss Program: Randomized Clinical Trial. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e14 [FREE Full text] [doi: 10.2196/mhealth.7713] [Medline: 29326093]
17. Kraschnewski JL, Stuckey HL, Rovniak LS, Lehman EB, Reddy M, Poger JM, et al. Efficacy of a weight-loss website based on positive deviance. A randomized trial. Am J Prev Med 2011 Dec;41(6):610-614. [doi: 10.1016/j.amepre.2011.08.012] [Medline: 22099238]
18. Krukowski RA, Tilford JM, Harvey-Berino J, West DS. Comparing behavioral weight loss modalities: incremental cost-effectiveness of an internet-based versus an in-person condition. Obesity (Silver Spring) 2011 Aug;19(8):1629-1635 [FREE Full text] [doi: 10.1038/oby.2010.341] [Medline: 21253001]
19. McConnon Á, Kirk SF, Cockroft JE, Harvey EL, Greenwood DC, Thomas JD, et al. The Internet for weight control in an obese sample: results of a randomised controlled trial. BMC Health Serv Res 2007 Dec 19;7:206 [FREE Full text] [doi: 10.1186/1472-6963-7-206] [Medline: 18093289]
20. Padwal RS, Klarenbach S, Sharma AM, Fradette M, Jelinski SE, Edwards A, et al. The evaluating self-management and educational support in severely obese patients awaiting multidisciplinary bariatric care (EVOLUTION) trial: principal results. BMC Med 2017 Mar 02;15(1):46 [FREE Full text] [doi: 10.1186/s12916-017-0808-6] [Medline: 28249576]
21. Steinberg DM, Tate DF, Bennett GG, Ennett S, Samuel-Hodge C, Ward DS. The efficacy of a daily self-weighing weight loss intervention using smart scales and e-mail. Obesity (Silver Spring) 2013 Sep;21(9):1789-1797 [FREE Full text] [doi: 10.1002/oby.20396] [Medline: 23512320]
22. Yardley L, Ware LJ, Smith ER, Williams S, Bradbury KJ, Arden-Close EJ, et al. Randomised controlled feasibility trial of a web-based weight management intervention with nurse support for obese patients in primary care. Int J Behav Nutr Phys Act 2014 May 21;11:67 [FREE Full text] [doi: 10.1186/1479-5868-11-67] [Medline: 24886516]
23. Pappa GL, Cunha TO, Bicalho PV, Ribeiro A, Couto Silva AP, Meira W, et al. Factors Associated With Weight Change in Online Weight Management Communities: A Case Study in the LoseIt Reddit Community. J Med Internet Res 2017 Dec 16;19(1):e17 [FREE Full text] [doi: 10.2196/jmir.5816] [Medline: 28093378]

Abbreviations

BMI: body mass index
GRADE: Grades of Recommendation, Assessment, Development and Evaluation
MD: mean difference

Edited by G Eysenbach; submitted 08.12.17; peer-reviewed by O Uthman, IN Gomez; comments to author 18.03.18; revised version received 14.08.18; accepted 30.08.18; published 08.01.19.

Please cite as:
Beleigoli AM, Andrade AQ, Cançado AG, Paulo MNL, Diniz MDFH, Ribeiro AL
Web-Based Digital Health Interventions for Weight Loss and Lifestyle Habit Changes in Overweight and Obese Adults: Systematic Review and Meta-Analysis
J Med Internet Res 2019;21(1):e298
URL: https://www.jmir.org/2019/1/e298/
doi:10.2196/jmir.9609
PMID:30622090

Review

Identifying Brief Message Content for Interventions Delivered via Mobile Devices to Improve Medication Adherence in People With Type 2 Diabetes Mellitus: A Rapid Systematic Review

Hannah Long[1], MSc; Yvonne K Bartlett[1], PhD; Andrew J Farmer[2], DM; David P French[1], PhD

[1]Manchester Centre for Health Psychology, School of Health Sciences, University of Manchester, Manchester, United Kingdom

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

Corresponding Author:
David P French, PhD
Manchester Centre for Health Psychology
School of Health Sciences
University of Manchester
Coupland 1 Building
Coupland Street
Manchester, M13 9PL
United Kingdom
Phone: 44 161 275 2605
Email: david.french@manchester.ac.uk

Abstract

Background: Current interventions to support medication adherence in people with type 2 diabetes are generally resource-intensive and ineffective. Brief messages, such as those delivered via short message service (SMS) systems, are increasingly used in digital health interventions to support adherence because they can be delivered on a wide scale and at low cost. The content of SMS text messages is a crucial intervention feature for promoting behavior change, but it is often unclear what the rationale is for chosen wording or any underlying mechanisms targeted for behavioral change. There is little guidance for developing and optimizing brief message content for use in mobile device–delivered interventions.

Objective: This review aimed to (1) identify theoretical constructs (ie, the targets that interventions aim to change) and behavioral strategies (ie, features of intervention content) found to be associated with medication adherence in patients with type 2 diabetes and (2) map these onto a standard taxonomy for behavior change techniques (BCTs, that is, *active ingredients* of interventions used to promote behavioral change, to produce an evidence-based set of approaches that have shown promise of improving adherence in previous studies and which could be further tested in digital health interventions.

Methods: A rapid systematic review of existing relevant systematic reviews was conducted. MEDLINE and PsycINFO databases were searched from inception to April 10, 2017. Inclusion criteria were (1) systematic reviews of quantitative data if the studies reviewed identified predictors of or correlates with medication adherence or evaluated medication adherence–enhancing interventions and included adult participants taking medication to manage a chronic physical health condition, and (2) systematic reviews of qualitative studies of experiences of medication adherence for adult participants with type 2 diabetes. Data were extracted on review characteristics and BCTs, theoretical constructs, or behavioral strategies associated with improved adherence. Constructs and strategies were mapped onto the BCT version 1 taxonomy.

Results: A total of 1701 references were identified; 25 systematic reviews (19 quantitative reviews, 3 qualitative reviews, and 3 mixed-method reviews) were included. Moreover, 20 theoretical constructs (eg, self-efficacy) and 19 behavioral strategies (eg, habit analysis) were identified in the included reviews. In total, 46 BCTs were identified as being related to medication adherence in type 2 diabetes (eg, habit formation, prompts or cues, and information about health consequences).

Conclusions: We identified 46 promising BCTs related to medication adherence in type 2 diabetes on which the content of brief messages delivered through mobile devices to improve adherence could be based. By using explicit systematic review methods and linking our findings to a standardized taxonomy of BCTs, we have described a novel approach for the development of digital message content. Future brief message interventions that aim to support medication adherence could incorporate the identified BCTs.

(J Med Internet Res 2019;21(1):e10421) doi:10.2196/10421

KEYWORDS

medication adherence; diabetes mellitus; systematic review; text messaging; mHealth; self-management

Introduction

Background

Diabetes mellitus affects an estimated 422 million people globally [1]. Approximately 90% of these cases are of type 2 diabetes [2,3]. Type 2 diabetes is typically managed through diet, physical activity, and oral medication. However, poor adherence to oral glucose-lowering medications in this population is common. The incidence of nonadherence is estimated to be between 38% and 93% [4-6], depending on the method used to define and measure adherence, and up to 37% of patients stop using oral glucose-lowering medications within 1 year of starting treatment [7]. This is a problem because medication nonadherence is associated with poorer clinical and health outcomes in patients with diabetes [8,9] and results in substantial financial burdens on Western health care systems because of higher rates of morbidity and mortality associated with the condition [10,11].

Effective interventions to increase medication adherence are needed to optimize health outcomes, quality of life, and cost-effective health care [12]. Despite the need, there is a lack of convincing evidence for interventions to improve patient adherence to prescribed medications for common chronic health problems. A Cochrane systematic review of 182 adherence-enhancing trials reported inconsistent effects on adherence, from highly heterogeneous interventions, and the majority of the lowest risk of bias trials (12 of 17) reported no improvements in adherence and clinical outcomes [13]. The interventions were generally multicomponent and complex, delivered by health care professionals and involving one-on-one education and counseling. Such interventions are time- and resource-intensive and, therefore, difficult to deliver to a large number of people or implement in usual practice settings. Similarly, complex interventions aimed at supporting medication adherence in patients with diabetes also report few improvements in adherence or clinical outcomes [14-18].

As of May 2014, 7 billion mobile phone subscriptions were in use worldwide [19]. In the United Kingdom, 65% to 92% of retired households own a mobile phone [20], and in 2017, 73% of adults accessed the internet through a smartphone or mobile device, which is double the rate recorded in 2011 [21]. The United States displays similar trends in increased mobile phone ownership; 77% and 95% of adults own a smartphone or mobile phone, respectively [22]. The increased use of mobile phone and modern technologies has created a new medium for digital health interventions delivered by mobile devices to support self-management of long-term conditions and potential solutions to problems such as medication nonadherence. Currently, the short message service (SMS) text message function of mobile phones offers a possible low-cost and wide-reaching means to deliver behavior change interventions. Brief messages delivered by SMS text message can be used in several ways: to send motivational and social support messages [23], to challenge maladaptive beliefs [24], or provide a cue to action [25]. This is important, as the impact on behaviors, such as medication adherence, is likely to differ depending on the message content.

A recent systematic review of brief messaging interventions for patients with type 2 diabetes concluded that these interventions may have the potential for supporting medication adherence [26]. However, the authors noted a widespread lack of explicit theoretical frameworks underpinning the included trials and called for greater use of theory in designing interventions to address the behavioral mechanisms through which changes to adherence may occur [26].

Behavior change theory can inform interventions by providing insight into factors that influence behavior, which can be targeted for behavioral change. This includes *theoretical constructs* (ie, those variables or constructs from theories that are targeted by interventions, eg, self-efficacy) and mechanisms underlying specific *behavioral strategies* (ie, techniques not necessarily linked to a single theory but incorporated in interventions because they predict behavior, eg, modeling). To translate these strategies into a standardized language, a comprehensive taxonomy of 93 *behavior change techniques* (BCTs) has been developed for use in behavior change interventions [27]. BCTs are discrete intervention components used to facilitate behavior change, such as problem solving and goal setting, and can be thought of as the *active ingredients* in interventions [27]. With this BCT taxonomy (BCTT), researchers can identify and evaluate specific techniques linked to intervention success, which facilitates the replication and optimization of existing or new interventions [27]. There is evidence that interventions that incorporate a greater number of BCTs tend to have larger effects on behavior than those incorporating a smaller number [28].

Grounding research in theory is crucial to progress understanding of which constructs need to change to, in turn, effectively change behavior. However, it is often unclear how message content has been developed, including what, if any, theory the messages are rooted in or how different types of messages may affect behavior. Without this, it is difficult to establish whether messages targeting constructs are effective or how to replicate these interventions. Toolkits for designing SMS-based interventions exist, but guidelines for developing the core message content are lacking (eg, The Center for Research in Implementation Science and Prevention [29]). An established process for developing brief message content, such as those delivered via SMS text message, is needed to optimize intervention effectiveness and facilitate evaluation of the most successful message types for changing behavior.

This review represents the first stage of a larger program of research to identify effective brief message content for use in interventions to support medication adherence in people with type 2 diabetes. We propose a novel approach to develop theory-based brief message content for use in interventions

supporting medication adherence for people with type 2 diabetes. Given the lack of effectiveness of existing interventions [13], there is a need to determine promising novel intervention content for the future. For this reason, this rapid systematic review considered the factors related to adherence in 2 stages. First, we considered the factors found to affect adherence across a broad range of physical health conditions in previous systematic reviews of quantitative research. Second, we considered systematic reviews of qualitative research of adherence, specifically in people with diabetes to better understand the psychosocial context of and influences on adherence in our target population.

Aims

The aims of this rapid review were to (1) systematically search for systematic reviews of quantitative studies that focused on medication adherence in physical health conditions and systematic reviews of qualitative studies that focused on adherence in type 2 diabetes, (2) extract data on theoretical constructs and behavioral strategies associated with medication adherence, and (3) map these constructs and strategies to BCTs to generate a list of BCTs that may show promise in improving medication adherence. We propose using these BCTs, to inform intervention targets and content of brief messages in future digitally based adherence-enhancing interventions.

Methods

Design

We conducted a rapid systematic review of systematic reviews. Rapid systematic reviews use methods to accelerate and streamline traditional systematic review processes while preserving the quality and rigor of review methods [30]. A rapid review is an appropriate method, given our aim to identify new potential intervention targets, rather than to conduct a comprehensive appraisal of the evidence base.

Search Strategy

A systematic search of 2 electronic databases (MEDLINE and PsycINFO) was conducted from database inception to April 10, 2017. Search terms for medication adherence were chosen by adapting Medical Subject Headings (MeSH) and keywords previously used [13]. Search strategies were reviewed by the team. The search terms used in MEDLINE were “medication adherence [MeSH]” OR “medication compliance” OR “patient compliance [MeSH]” AND “review literature as topic [MeSH]” OR “systematic review” OR “meta-analysis” OR “meta-synthesis.” The search terms used in PsycINFO were “treatment compliance [MeSH]” OR “compliance [MeSH]” AND “literature review [MeSH]” OR “meta-analysis [MeSH]” OR “systematic review” OR “meta-synthesis.” EndNote reference software (Clarivate Analytics) was used to organize references.

Inclusion and Exclusion Criteria

Scope of Included Reviews

Systematic reviews of quantitative data, qualitative data, or mixed-method data (reporting quantitative and qualitative data) were included. Only papers written in English were included. Systematic reviews of adult (aged older than 18 years) samples taking medication to self-manage a diagnosed physical health condition were included. Review samples with children and adolescents (aged less than 18 years) were excluded, as were samples from exclusively nondeveloped countries.

Systematic reviews of quantitative studies had to report on (1) interventions to improve patients’ medication adherence or (2) predictors of or correlates with medication adherence in physical health conditions. Systematic reviews had to report subanalyses of medication adherence if adherence was not the primary outcome measure. Reviews had to include a behavioral measure of medication adherence or a combination of behavioral and clinical measures. Reviews that reported only clinical measures of adherence were excluded, as these measures may be affected by behaviors besides medication adherence. Reviews had to report data on patient adherence in chronic physical health conditions. We excluded reviews of patient samples with severe mental health or psychiatric conditions, substance abuse, acute-only conditions, contraceptive or sexual function medication, herbal remedies, vitamins, vaccinations, homeopathy, or conditions that could be chronic (eg, cancer) but for which the treatment medication is short-term (eg, oral chemotherapy). Reviews of acute physical health conditions and nonsevere mental health conditions (eg, diabetes with comorbid depression or depressive symptoms) were eligible if they also reviewed chronic conditions.

Systematic reviews of qualitative studies of adult samples in which diabetes was the main health condition were included. Participants with other physical health conditions were eligible if diabetes was the majority health condition in the sample. To be eligible, mixed-method systematic reviews had to report findings in line with the above criteria for either the quantitative reviews or the qualitative reviews.

Theoretical Constructs and Behavioral Strategies

To identify individual theoretical constructs or behavioral strategies that were associated with medication adherence, reviews of quantitative studies had to report distinct constructs or strategies and their relationship with medication adherence through pooling data across studies. Reviews of quantitative studies were excluded if they did not breakdown interventions and studies into the separate components and examine the effects. However, reviews of quantitative studies that analyzed similar behavioral strategies under an umbrella term were included and assessed as a single strategy (eg, *self-monitoring of medication adherence* using different methods of self-monitoring). Reviews of quantitative studies were excluded if they examined only the effects of (1) the form of the interventions or studies (eg, intervention duration and interventionist), (2) sociodemographic variables (eg, ethnicity and gender), or (3) different lifestyle behaviors (eg, dietary intake and alcohol consumption) on medication adherence.

Systematic reviews of qualitative studies had to report patients’ experiences of, or barriers and facilitators to, taking and adhering to medication for type 2 diabetes.

Data Extraction

Separate data extraction sheets were developed, piloted, and refined for quantitative and qualitative data. First, we extracted data from the systematic reviews of quantitative studies on a number of indicators (Multimedia Appendix 1) and the relationship between constructs and strategies with medication adherence.

Second, we extracted qualitative data in the form of the authors' narrative results sections. Results pertaining only to patients' experiences of adherence were extracted. The mixed-method reviews, and the data reported in these, were combined with either the pool of quantitative or qualitative reviews depending on whether the findings were more quantitative or qualitative in nature.

Identifying and Mapping Theoretical Constructs and Behavioral Strategies to Applicable Behavior Change Techniques

There were 2 stages involved in the process of identifying and mapping theoretical constructs and behavioral strategies to BCTs. First, we used the reviews of quantitative research to identify foundational constructs and strategies that were positively related to adherence in a range of physical health conditions. For each of the constructs and strategies identified in the reviews of quantitative studies, a judgment was made about which BCTs in the BCTT may be an applicable technique to change that particular construct or strategy. This process was completed iteratively, reviewing and refining judgments to ensure that all constructs and strategies had been paired with at least one BCT, and any relevant BCTs were captured. The process of mapping involved a degree of brainstorming [31] and idea generation, a recognized early step in intervention mapping, with *the aim of generating as many explanations as possible in response to a question* [32]. We have adapted this approach, generating *answers*, in the form of BCTs, to the constructs and strategies (the *questions*) extracted from the reviews.

Second, we took a diabetes-specific approach and used the reviews of qualitative research to consider the specific context of our target population and to ensure that we did not overlook constructs or strategies that were particularly relevant to adherence in patients with diabetes. In this stage, a different method was used to identify constructs and strategies in the qualitative data and link these to BCTs. The reviews of quantitative studies, by nature, were generally explicit in the reporting of constructs or strategies, but these were implicit in the qualitative review data. Given this, to identify which constructs or strategies may be related to patient adherence, a reviewer first familiarized herself by reading and rereading the extracted data. Inductive line-by-line coding was conducted, taking into consideration the meaning and context of different barriers and facilitators to adherence. The codes applied were the constructs and strategies identified from the quantitative findings or concepts from relevant psychological theory. The reviewer then followed the same process of mapping to BCTs as that used for the quantitative review data.

The initial findings and mapped BCTs from the quantitative and qualitative data were discussed within the research team to provide a sense check and to draw from additional expertise. In light of this, the first author further refined the associations between constructs and strategies and BCTs.

Results

Search Results

The electronic database searches identified 1701 references. After duplicates were removed, 1507 references remained. A reviewer screened all title or abstracts, and a second reviewer screened 29.86% (450/1507) of these (84% agreement). Moreover, 1296 references were excluded at the title-abstract stage. A reviewer read the full texts of the remaining 211 references, and the second reviewer read 29.9% (63/211) of these (71% agreement). Consensus on eligibility at both the title-abstract and full-text stage was reached through discussion. Any disagreements were resolved by discussing proposed reasons for inclusion and exclusion against the eligibility criteria. A total of 25 eligible systematic reviews (19 quantitative reviews, 3 qualitative reviews, and 3 mixed-methods reviews) published between 1995 and 2017 were included in this review—23 systematic reviews from database searches and 2 additional systematic reviews known to the researchers (Figure 1 illustrates the study inclusion process). From this point, 2 mixed-method reviews [33,34] have been integrated with the reviews of quantitative studies and the third mixed-method review [35] has been integrated with the reviews of qualitative studies.

Systematic Review Characteristics

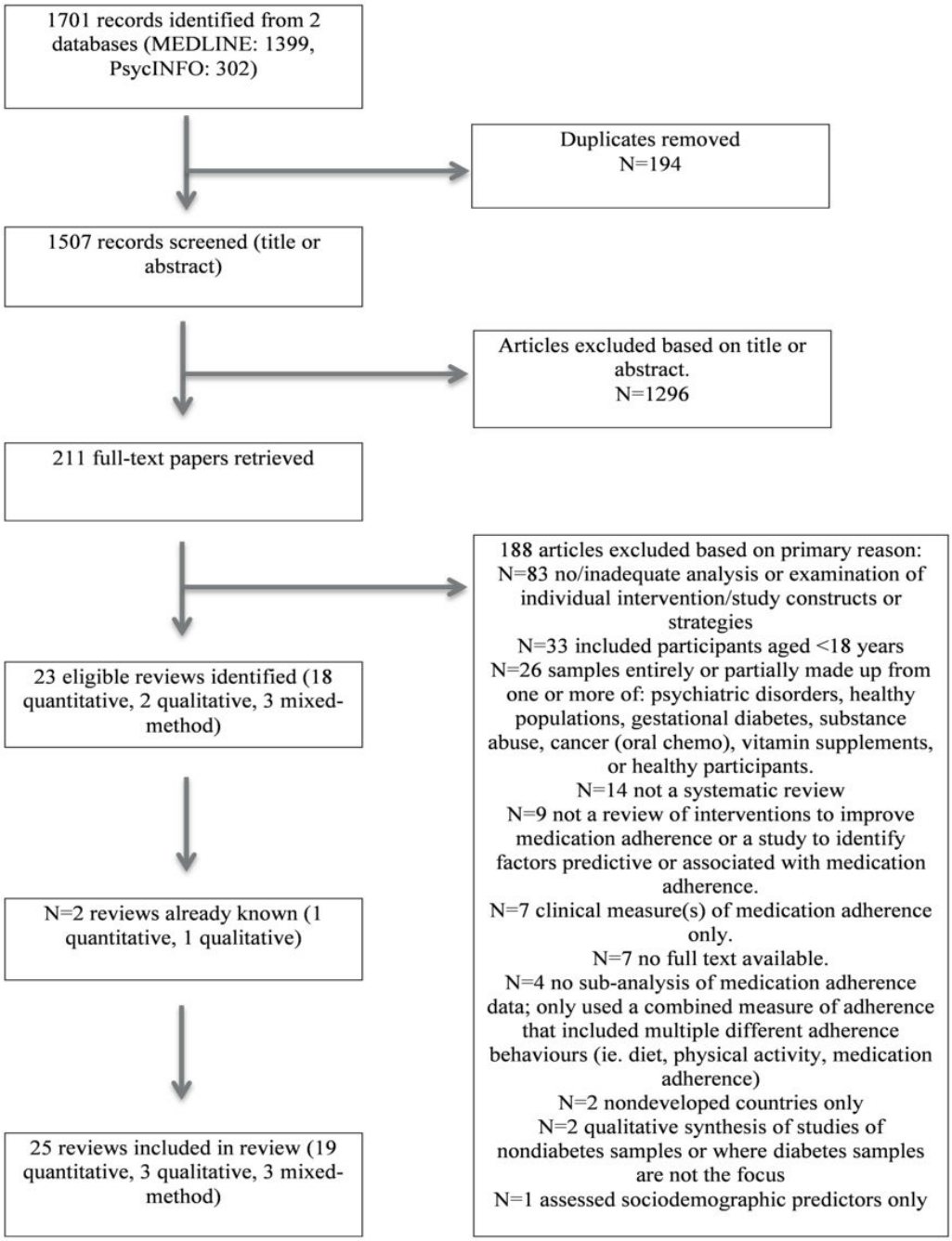
Systematic Review Aims

The reviews of quantitative studies examined interventions to promote medication adherence (k=16), predictive studies, or correlates of medication adherence (k=4) or both (k=1). The reviews of quantitative studies identified a total number of 1930 studies. The reviews of qualitative studies examined patients' understanding of barriers and facilitators to adherence in type 2 diabetes (k=2), medication adherence–related beliefs and decision-making processes regarding taking medication in diabetes and cardiovascular disease patients (k=1), and patients' perceptions and experiences of taking oral medication for type 2 diabetes (k=1). The reviews of qualitative studies identified a total number of 140 studies.

Patient Populations

Of the reviews of quantitative studies, 5 examined medication adherence in multiple chronic conditions and 4 reviews were of multiple chronic and acute conditions. The remaining reviews examined adherence in single conditions: hypertension (k=3), HIV and AIDS (k=2), type 2 diabetes (k=2), cardiovascular disease (k=1), chronic pain (k=1), heart failure (k=1), psoriasis (k=1), and organ transplantation (k=1). Overall, 2 reviews examined interventions targeted at health care providers to promote patient medication adherence. In addition, 3 reviews of qualitative studies included samples of diabetes patients, and the fourth review included a combination of samples of diabetes and cardiovascular disease patients.

Figure 1. Flow diagram of study inclusion and exclusion process.



Characteristics of Primary Studies

The primary studies in the reviews of quantitative studies were of various study designs; 14 reviews assessed multiple primary study designs. The study designs (when reported by review authors) included randomized controlled trials (RCTs; k=17), nonrandomized (k=8), pre- and posttest (k=4), prospective correlational (k=6), and cross-sectional correlational (k=5) designs (see Multimedia Appendix 1 for a summary table of review characteristics).

Quantitative Systematic Review Findings

In total, 14 theoretical constructs and 13 behavioral strategies were extracted from 18 of 21 systematic reviews of quantitative studies because they were related to medication adherence (Table 1). There was variation in the methods used by review authors to analyze the relationship of a construct or strategy to medication adherence. The different approaches to analysis included moderation analysis, correlation analysis, calculating a tally of the number of primary studies that had found a statistically significant effect when a particular construct or strategy was present and, where possible, associated effect sizes, and qualitative comparative analysis to determine the necessity or sufficiency of an individual construct or strategy to influence medication adherence.

The 14 theoretical constructs and 13 behavioral strategies were mapped to 34 BCTs (Table 2). Each construct or strategy was linked with between 1 and 13 associated BCTs.

Table 1. Theoretical constructs and behavioral strategies associated with improved medication adherence extracted from the quantitative reviews.

Authors (year)	TC[a] and BS[b]	Evidence summary
Broekmans et al (2009) [33]	• Medication-related concerns (TC) • Poor patient-physician communication and satisfaction (TC)	• Concerns about side effects were a significant[c] correlate of lower MA[d] in k=1. Fewer concerns about withdrawal were a significant[c] correlate of lower MA (k=1). • Poor patient-physician communication and satisfaction were a significant[c] correlate of lower MA (k=1).
Conn et al (2009) [36]	• Coping with side effects (BS) • Stimulus to take medication (BS) • Self-monitoring of symptoms related to medications (BS) • Providing succinct written instructions (BS)	• Adding *side-effect management* as a moderator to a multiple-moderator model significantly[e] improved the model (beta=.60). • Interventions, including a stimulus to take medication were more effective at improving MA (ES[f] 1.06) than interventions without these cues (ES 0.30). • Interventions that directed participants to self-monitor symptoms related to medications (including symptom improvement from taking medications and medication side effects) were more effective (ES 1.18) at improving MA than interventions that lacked this component (ES 0.30). • Interventions with succinct written instructions achieved better effects on MA (ES 0.61) than studies without succinct written instructions (ES 0.29).
Conn et al (2015a) [37]	• No TC or BS associated with MA	• N/A[g]
Conn et al (2015b) [38]	• No TC or BS associated with MA	• N/A
Conn et al (2016) [39]	• Habit analysis (BS) • Prompts or cues (BS)	• Habit-focused interventions in which participants' daily habits were linked to taking medications were more effective at increasing MA relative to interventions that lacked this component (0.57 vs 0.22[e]). • Studies that used prompts or cues for taking medications had larger ES than studies that did not (0.50 vs 0.23[c]).
Conn et al (2017) [40]	• Habit analysis (BS)	• Moderation analysis showed that interventions that included habit analysis were more effective (*d*=0.37) at improving MA than interventions that did not (*d*=0.28).
Cutrona et al (2010) [41]	• Reinforcement and reminding (BS)	• The majority of k=16 showed small effects of reinforcement and reminding on MA, whereas k=3 yielded large effects.
Devine et al (1995) [42]	• Self-monitoring of medications (BS) • Self-monitoring of symptoms related to medications (BS) • Increasing health-related knowledge through education (BS)	• Effect size values on MA by type of treatment were monitoring medications (*d*=0.43), monitoring blood pressure (*d*=0.37), and education (*d*=0.81).
Dew et al (2007) [43]	• Social support (BS)	• Poorer social support was significantly associated with greater nonadherence (ES 0.10, CI 0.03-0.26[c]) from k=11.
Farmer et al (2015) [26]	• Self-monitoring of medications (BS)	• Overall, 3 of 6 self-monitoring trials observed significant[c] improvements in MA.
Fogarty et al (2002) [34]	• Social support (BS) • Scheduling demands (TC) • Regimen complexity (TC)	• Social support was statistically significantly associated with MA in 1 of 4 papers and 1 of 8 abstracts. • A total of 2 of 15 abstracts and 2 of 5 papers reported a significant[c] association between scheduling demands and MA. • Overall, 2 of 17 abstracts and 3 of 4 studies reported a significant association between regimen complexity and MA. In total, 2 of 4 of these found the direction of the association to be as expected; more complex regimens were associated with decreased MA.

Authors (year)	TC[a] and BS[b]	Evidence summary
Holmes et al (2014) [44]	• Self-efficacy (TC) • Perceived barriers (TC) • Perceived adverse effects (TC) • Perceived benefits (TC) • Perceived severity (TC) • Perceived susceptibility (TC) • Attitude (TC) • Intention (TC) • Perceived behavioral control (TC) • Necessity beliefs (TC) • Medication-related concerns (TC)	• Self-efficacy was a significant[c] predictor of MA in 7 of 7 studies of sociocognitive theory, 6 of 6 studies of self-regulation theory, and 4 of 6 studies of social support theory. • Perceived barriers were significantly[c] associated with MA in 11 of 17 studies. • Perceived adverse effects were significantly[c] associated with MA in 4 of 5 studies. • Perceived benefits were significantly[c] associated with MA in 5 of 11 studies. • Perceived severity was significantly[c] associated with MA in 3 of 7 studies. • Perceived susceptibility was significantly[c] associated with MA in 3 of 6 studies. • Attitude was significantly[c] associated with MA in 2 of 5 studies. • Intention was significantly[c] associated with MA in 2 of 5 studies. • Perceived behavioral control was significantly[c] associated with MA in 2 of 4 studies. • Necessity beliefs were significantly[c] associated with MA in 7 of 8 studies. • Medication-related concerns were significantly[c] associated with MA in 7 of 8 studies.
Kahwati et al (2016) [45]	• Self-efficacy (TC) • Attitude (TC) • Increasing health-related knowledge through education (BS) • Motivational interviewing (BS)	• Enhancing self-efficacy was identified as individually sufficient for improving MA (consistency 90%). • Improving attitude was identified as individually sufficient for improving MA (consistency 90%). • Increasing knowledge was a necessary individual BCT for improved MA; it was present in 31 of 34 studies (consistency 91%). • Motivational interviewing was identified as close to the consistency threshold for an individually sufficient technique for improving MA (consistency 78%).
Ruppar et al (2015) [46]	• No TC or BS associated with MA	• N/A
Schedlbauer et al (2010) [47]	• Reinforcement and reminding (BS)	• In total, 4 of 6 studies reported statistically[c] improved MA following reminders in the form of written postal material (k=1), regular telephone calls (k=2), and a simple calendar reminder of medication taking (k=1).
Simoni et al (2006) [48]	• Interactive discussion of cognitions, motivations, and expectations about adherence (BS)	• Interactive discussion of cognitions, motivations, and expectations about MA, ES 1.62 (CI 1.21-2.03; k=14) versus no discussion ES 0.99 (CI 0.55-1.79; k=4).
Takiya et al (2004) [49]	• Prompts or cues (BS) • Increasing health-related knowledge through education (BS)	• Beeper: 1 of 1 study reported significant improvement in MA, ES 0.09 (CI −0.15 to 0.31[c]). • Phone reminder: 1 of 1 study reported significant improvement in MA, ES 0.03 (CI −0.09 to 0.15[c]). • Increasing health-related knowledge through education: 2 of 3 studies reported significant improvement in MA, ES 0.18 (CI −0.11 to 0.44[c]), 0.03 (CI −0.26 to 0.30[c]).
Teeter et al (2014) [50]	• Motivational interviewing (BS)	• Overall, 6 of 9 studies reported statistically significant[c] differences between intervention and control groups for change in MA.
Thorneloe et al (2013) [51]	• Patient satisfaction with their treatment (TC)	• k=1 reported patients being too busy or fed up was associated with reduced MA.
Xu et al (2014) [52]	• Tailoring care plan (BS)	• Tailoring was the most common persuasive attribute; 76% of interventions that successfully improved MA included tailoring versus 33% of interventions in which MA did not improve[e] (the number of included studies that incorporated tailoring was not reported).

Authors (year)	TC[a] and BS[b]	Evidence summary
Zomahoun et al (2015) [53]	• Coping with side effects (BS)	• Interventions in which *cope with side effects* was applied had a pooled SMD[h] of 0.64 (95% CI 0.31-0.96) versus 0.02 (95% CI −0.25 to 0.28) for those who did not (the subgroup differences[e]).

[a]TC: theoretical construct.
[b]BS: behavioral strategy.
[c]*P*<.05.
[d]MA: medication adherence.
[e]*P*<.01.
[f]ES: effect size.
[g]N/A: not applicable.
[h]SMD: standard mean difference.

Table 2. Theoretical constructs and behavioral strategies identified by the reviews of quantitative studies and behavior change techniques mapped to these by the research team.

Theoretical constructs and behavioral strategies	Behavior change techniques
Theoretical constructs	
Attitude	Framing or reframing, pros and cons, information about emotional consequences, information about health consequences, information about social and environmental consequences, and salience of consequences.
Intention	Anticipated regret, comparative imagining of future outcomes, pros and cons, and verbal persuasion about capability.
Medication-related concerns	Comparative imagining of future outcomes, framing or reframing, information about emotional consequences, information about health consequences, information about social and environmental consequences, problem solving, prompts or cues, pros and cons, reduce negative emotions, salience of consequences, social support - emotional, social support - practical, and social support - unspecified.
Necessity beliefs	Anticipated regret, comparative imagining of future outcomes, framing or reframing, information about emotional consequences, information about health consequences, information about social and environmental consequences, pros and cons, and salience of consequences.
Patient satisfaction with their treatment	Framing or reframing, pros and cons, and reduce negative emotions.
Perceived adverse effects	Comparative imagining of future outcomes, framing or reframing, and incompatible beliefs.
Perceived barriers	Framing or reframing, pros and cons, restructuring the physical environment, restructuring the social environment, social support - emotional, social support - practical, and social support - unspecified.
Perceived behavioral control	Focus on past success, mental rehearsal of successful performance, self-talk, and verbal persuasion about capability.
Perceived benefits	Anticipated regret, comparative imagining of future outcomes, information about emotional consequences, information about health consequences, information about social and environmental consequences, pros and cons, and salience of consequences.
Perceived severity	Anticipated regret, comparative imagining of future outcomes, feedback on outcomes of behavior, framing or reframing, incompatible beliefs, information about emotional consequences, information about health consequences, and information about social and environmental consequences.
Perceived susceptibility	Comparative imagining of future outcomes, framing or reframing, information about health consequences, and pros and cons.
Regimen complexity	Habit formation, problem solving, and prompts or cues.
Scheduling demands	Action planning and problem solving.
Self-efficacy	Focus on past success, identification of self as role model, mental rehearsal of successful performance, self-talk, social reward, valued self-identity, and verbal persuasion about capability.
Behavioral strategies	
Coping with side effects	Problem solving, social support - emotional, social support - practical, and social support - unspecified.
Habit analysis	Behavioral practice or rehearsal, habit formation, habit reversal, and graded tasks.
Increasing health-related knowledge through education	Information about antecedents, information about emotional consequences, information about health consequences, information about social and environmental consequences, and instruction on how to perform a behavior.
Interactive discussion of cognitions and motivations and expectations about adherence	Anticipated regret, comparative imagining of future outcomes, framing or reframing, and pros or cons.
Motivational interviewing	Comparative imagining of future outcomes, framing or reframing, pros or cons, and social support (emotional).
Prompts and cues	Prompts or cues.
Providing succinct written instructions	Instruction on how to perform a behavior.
Reinforcement and reminding	Behavioral practice or rehearsal, habit formation, and prompts or cues.
Self-monitoring of medications	Self-monitoring of behavior.
Self-monitoring of symptoms related to medications	Self-monitoring of outcome(s) of behavior.
Social support	Social support - emotional, social support - practical, and social support - unspecified.

Theoretical constructs and behavioral strategies	Behavior change techniques
Stimulus to take medication	Prompts and cues.
Tailoring care plan	Information about health consequences and social support (emotional).

Qualitative Systematic Review Findings

In total, 17 theoretical constructs and 12 behavioral strategies were coded in the qualitative data [35,54-56] (Multimedia Appendix 2). Of these, 11 were not identified by the reviews of quantitative data, including 6 constructs: identity, motivation, negative emotions, response efficacy, social comparison, and social context and 6 strategies: credible source, demonstration of the behavior, problem solving, self-adjustment and experimentation with medication dose and frequency, self-management strategies, and self-monitoring of outcome(s) of behavior). The 17 constructs and 12 strategies were mapped to 46 BCTs; 12 BCTs were identified in addition to the 34 BCTs mapped from the quantitative review findings. Each construct or strategy had between 1 and 25 associated BCTs (Table 3).

Selection of Candidate Behavior Change Techniques

In total, a pool of 46 BCTs were mapped from the theoretical constructs and behavioral strategies identified in the included reviews (Textbox 1, see Multimedia Appendix 3 for BCT definitions from the study by Michie et al [27]).

Table 3. Theoretical constructs and behavioral strategies identified in the reviews of qualitative studies and the mapped behavior change techniques from the behavior change technique taxonomy.

Theoretical constructs and behavioral strategies	Behavior change techniques
Theoretical construct	
Attitude	Anticipated regret, framing or reframing, information about health consequences, and pros and cons.
Identity	Identity associated with changed behavior, identification of self as a role model, incompatible beliefs, and valued self-identity.
Medication-related concerns	Adding objects to the environment, anticipated regret, behavioral experiments, comparative imagining of future outcomes, conserving mental resources, credible source, demonstration of behavior, feedback on outcome(s) of behavior, framing or reframing, habit formation, information about emotional consequences, information about health consequences, information about social and environmental consequences, instruction on how to perform the behavior, monitoring of emotional consequences, problem solving, prompts or cues, pros and cons, reduce negative emotions, restructuring the physical environment, salience of consequences, self-monitoring of outcome(s) of behavior, social support - emotional, social support - practical, and social support - unspecified.
Motivation	Anticipated regret, framing or reframing, salience of consequences, and self-talk.
Necessity beliefs	Anticipated regret, behavioral practice or rehearsal, comparative imagining of future outcomes, feedback on outcome(s) of behavior, framing or reframing, information about emotional consequences, information about health consequences, information about social and environmental consequences, habit formation, pros and cons, salience of consequences, self-monitoring of behavior, and self-monitoring of outcome(s) of behavior.
Negative emotions	Framing or reframing, information about emotional consequences, monitoring of emotional consequences, reattribution, reducing negative emotions, social support - emotional, and verbal persuasion about capability.
Patient-physician relationship and communication	Credible source, framing or reframing, information about health consequences, social support - emotional, and social support - practical.
Perceived barriers	Avoidance or reducing exposure to cues for the behavior, conserving mental resources, credible source, demonstration of the behavior, framing or reframing, habit formation, identification of self as a role model, information about antecedents, information about emotional consequences, information about health consequences, information about others' approval, information about social and environmental consequences, instruction on how to perform the behavior, problem solving, prompts or cues, reducing negative emotions, restructuring the social environment, salience of consequences, social support - practical, social support - unspecified, and valued self-identity.
Perceived behavioral control	Anticipated regret, behavioral practice or rehearsal, focus on past success, framing or reframing, information about antecedents, mental rehearsal of successful performance, reattribution, social support - unspecified, and verbal persuasion about capability.
Perceived benefits	Anticipated regret, comparative imagining of future outcomes, feedback on outcomes of behavior, framing or reframing, incompatible beliefs, information about emotional consequences, information about health consequences, information about others' approval, information about social and environmental consequences, and pros and cons.
Perceived seriousness	Anticipated regret, comparative imagining of future outcomes, feedback on outcome(s) of behavior, framing or reframing, information about emotional consequences, information about health consequences, information about others' approval, information about social and environmental consequences, and pros and cons.
Perceived susceptibility	Anticipated regret, comparative imagining of future outcomes, information about health consequences, pros and cons, and reattribution.
Regimen complexity	Conserving mental resources, habit formation, problem solving, and prompts or cues.
Response efficacy	Anticipated regret, credible source, feedback on outcome(s) of behavior, information about health consequences, pros and cons, and self-monitoring of outcome(s) of behavior.
Self-efficacy	Behavioral practice or rehearsal, feedback on outcome(s) of behavior, focus on past success, graded tasks, identification of self as a role model, information about others' approval, mental rehearsal of successful performance, monitoring of emotional consequences, reduce negative emotions, self-talk, social reward, social support - emotional, and verbal persuasion about capability.
Social comparison	Anticipated regret, comparative imagining of future outcomes, information about others' approval, social comparison, social support - emotional, social support - practical, and social support -unspecified.

Theoretical constructs and behavioral strategies	Behavior change techniques
Social context (support, influence, and stigma)	Avoidance or reducing exposure to cues for the behavior, credible source, demonstration of the behavior, generalization of a target behavior, identification of self as role model, incompatible beliefs, information about antecedents, information about health consequences, information about others' approval, restructuring the social environment, social comparison, social support - emotional, social support - practical, social support - unspecified, and valued self-identity.
Behavioral strategy	
Coping with side effects	Anticipated regret, information about health consequences, problem solving, social support - emotional, social support - practical, and reattribution.
Credible source	Credible source.
Demonstration of the behavior	Demonstration of the behavior.
Habits	Action planning, behavioral practice or rehearsal, generalization of target behavior, graded tasks, habit formation, and habit reversal.
Health-related information and knowledge	Action planning, credible source, information about emotional consequences, information about health consequences, information about social and environmental consequences, instruction on how to perform the behavior, reattribution, salience of consequences, and social support - practical.
Problem solving	Action planning and problem solving.
Prompts and reminders	Adding objects to the environment, prompts or cues, and restructuring the physical environment.
Self-adjustment and experimentation with medication dose and frequency	Anticipated regret, behavioral experiments, comparative imagining of future outcomes, framing or reframing, generalization of a target behavior, information about health consequences, reattribution, problem solving, pros and cons, self-monitoring of behavior, and self-monitoring of outcome(s) of behavior.
Self-management strategies	Behavioral practice or rehearsal, generalization of a target behavior, graded tasks, and habit formation.
Self-monitoring of outcome(s) of behavior	Behavioral experiments, feedback on outcome(s) of behavior, self-monitoring of behavior, and self-monitoring of outcome(s) of behavior.
Self-monitoring of symptoms	Information about health consequences, self-monitoring of behavior, and self-monitoring of outcome(s) of behavior.
Tailoring care plan	Action planning, information about health consequences, graded tasks, and problem solving.

Textbox 1. The 46 behavior change techniques (grouped according to the behavior change technique taxonomy version 1) identified from systematic reviews of quantitative and qualitative studies as being promising candidates for future brief message interventions.

1. Goals and planning

1.2. Problem solving

1.4. Action planning

2. Feedback and monitoring

2.3. Self-monitoring of behavior

2.4. Self-monitoring of outcome(s) of behavior

2.7. Feedback on outcome(s) of behavior

3. Social support

3.1. Social support (unspecified)

3.2. Social support (practical)

3.3. Social support (emotional)

4. Shaping knowledge

4.1. Instruction on ow to perform the behavior

4.2. Information about antecedents

4.3. Reattribution

4.4. Behavioral experiments

5. Natural consequences

5.1. Information about health consequences

5.3. Information about social and environmental consequences

5.2. Salience of consequences

5.4. Monitoring of emotional consequences

5.5. Anticipated regret

5.6. Information about emotional consequences

6. Comparison of behavior

6.1. Demonstration of behavior

6.2. Social comparison

6.3. Information about others' approval

7. Associations

7.1. Prompts or cues

8. Repetition and substitution

8.1. Behavioral practice or rehearsal

8.3. Habit formation

8.6. Generalization of target behavior

8.4. Habit reversal

8.7. Graded tasks

9. Comparison of outcomes

9.1. Credible source

9.2. Pros and cons

9.3. Comparative imagining of future outcomes

10. Reward and threat

10.4. Social reward

11. Regulation

11.2. Reduce negative emotions
11.3. Conserving mental resources
12. Antecedents
12.1. Restructuring the physical environment
12.2. Restructuring the social environment
12.3. Avoidance or reducing exposure to cues for the behavior
12.5. Adding objects to the environment
13. Identity
13.1. Identification of self as role model
13.2. Framing or reframing
13.3. Incompatible beliefs
13.4. Valued self-identity
13.5. Identity associated with changed behavior
14. Self-belief
14.1. Verbal persuasion about capability
14.2. Mental rehearsal of successful performance
14.3. Focus on past success
14.4. Self-talk

Discussion

Principal Findings

This rapid review identified 25 published systematic reviews of medication adherence in patients with physical health conditions and extracted a total of 20 theoretical constructs and 19 behavioral strategies associated with medication adherence. These constructs and strategies were mapped to 46 applicable BCTs from the BCTT version 1 [27], which can be used in future medication adherence interventions. In the first stage, the reviews of quantitative research gave rise to 34 BCTs related to adherence in a broad range of chronic physical health conditions. To ensure the specific context of adherence in type 2 diabetes was accounted for, 12 additional BCTs were identified following review of the diabetes-specific qualitative data.

Strengths

This review has several strengths. It is the first rapid review of systematic reviews that aimed to identify theoretical constructs and behavioral strategies related to medication adherence in patients with type 2 diabetes. We have followed established and explicit systematic review methodology for this rapid review; thus, minimizing bias in literature searching, retrieval, and appraisal [57]. Furthermore, taking a rapid systematic approach to synthesize an evidence base and using systematic review methods to consider existing systematic reviews are 2 relatively new approaches to evidence synthesis. In addition, this review proposes a novel approach to develop message content, which could be used to inform future brief message–based or other electronic health and mobile health (mHealth) interventions. Use of the methodology of this review may lead to the identification of promising BCTs for promoting alternative health behaviors, as it is likely that different BCTs will be relevant to different contexts. This has produced findings in line with the validated BCTT [27], a leading classification system by behavior change researchers. Finally, the use of theory is often missing or inadequately reported in interventions for enhancing adherence in patients with type 2 diabetes, which makes it difficult to interpret why, how, and where theory may have impacted intervention success [26]. This rapid review starts to address these issues, proposing why, how, and where to incorporate BCTs in brief message intervention design.

Limitations

This review has several limitations. We used a simple search strategy to facilitate a rapid review. Although this was appropriate for our aims, forward and backward citation searching may have improved the comprehensiveness of the literature identified. A number of systematic reviews were excluded because the independent relationship between theoretical constructs and behavioral strategies and medication adherence were not reported, and instead, the focus was on establishing the overall effectiveness of generally complex interventions. The quality of the included reviews was not assessed and, therefore, the possible risk of bias in the included systemic reviews is unknown. However, eligible reviews had to have been conducted systematically to maintain a minimum requirement of methodological rigor. Furthermore, consensus on the definition, and essential methodological processes, of rapid reviews has not been reached [58-60] and the necessity of the quality assessment stage has been debated [30]. We adhered to a set of defining characteristics proposed for rapid reviews by recent research [30]. In a modified Delphi consensus study, and based on the opinion of 66 literature experts, Kelly and colleagues [30] concluded that rapid reviews (1) are conducted in a shorter time frame than systematic reviews

(completed in up to 3 months, compared with an average 15 months for a full systematic review) [61-63], (2) use the most systematic and rigorous methods to synthesize evidence and answer the research question(s) as the time limit permits, (3) tailor methods typical of a systematic review to accelerate the review process, and (4) are transparent in reporting all methods and findings. By following these guidelines, we have produced a methodologically strong rapid review.

Our method of mapping was adapted from the intervention mapping approach and incorporated the subjective views of the authors. In future, primary research may benefit from using more rigorous and reproducible methods of mapping. This may be achieved through the use of established coding frames and mapping techniques derived more heavily from the intervention mapping literature. Valuable research in the area of mapping theoretical constructs to BCTs is underway and this should serve to progress and inform future endeavors of a similar nature [64]. However, there is currently no consensus on how to do this [64] and, as such, our approach was appropriate for our aim to identify a selection of BCTs that can be used as a basis for future research.

Implications and Future Research

The primary practical contribution of this review is the identification of promising intervention content in the form of a set of 46 BCTs. These could be used by researchers to form the basis for developing brief messages in interventions to promote medication adherence in people with type 2 diabetes. Such interventions are a promising avenue for improving adherence [26]. Given the aims of our wider program of research, we advocate this particular application of our findings, but our findings are not limited to brief messages or populations with type 2 diabetes. The comprehensive list of BCTs could be incorporated into other modes of intervention delivery, with a variety of interactive capabilities, for example, mobile phone apps. In addition, the 34 BCTs mapped solely from the quantitative review findings may be applied more broadly to medication adherence in chronic physical health conditions, provided unique social and contextual influences on adherence were taken into account.

The primary research implication of this review is that it facilitates the development of a body of brief messages based on the identified BCTs. Each message could incorporate a BCT and use it to frame and support different aspects of adherence behavior in people with type 2 diabetes. It will be important to ensure that messages accurately reflect the BCTs with which they are associated. The process of content development may be improved by seeking expert input. This is the next planned step in our program of research; health care professionals and behavioral scientists will be asked to generate brief messages based on the target BCTs. These messages and associated BCTs will be further tested in people with type 2 diabetes to assess credibility and acceptability, and subsequently in behavioral scientists to assess fidelity. Following this, a feasibility trial and a full RCT will assess the effectiveness of the brief message intervention to support medication adherence in people with type 2 diabetes. The approach of this review may be particularly suited to the area of adherence, given that the need for new approaches has been identified by leading systematic reviews [13]. More broadly, other researchers may wish to adapt our approach to develop novel interventions for other behaviors or populations where there is a need.

This review is a starting point; we have cast a wide net to suggest new intervention content, given that existing interventions have been found to have little effect on medication adherence. Further research is needed to establish more precisely which BCTs might be most effective at supporting medication adherence in particular contexts as well as the factors that may influence effectiveness. In the context of diabetes, these factors could include health literacy, diabetes health competency, and numeracy. This is complex to establish, and such questions were beyond the scope of this review. When selecting BCTs, researchers should be guided by what they understand of the context and mode of intervention delivery and nominate BCTs that suit their particular setting, population, and research aims while considering any relevant disease-specific research. It will be important to achieve a balance between learning from previous evidence in the field and testing novel approaches to support adherence.

Conclusions

This review has identified a range of theoretical constructs and behavioral strategies related to medication adherence and mapped these to 46 BCTs that may show promise in supporting adherence in people with type 2 diabetes. We propose developing and testing the effectiveness of brief message content in SMS text messages based on the 46 BCTs and have, thus, described a novel approach to designing the content of brief messages to optimize mHealth interventions.

Acknowledgments

This review presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research program (RP-PG-1214-20003). The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health and Social Care. The SuMMiT-D research team acknowledges the support of the NIHR through the Clinical Research Network.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Main characteristics of the included systematic reviews.

[PDF File (Adobe PDF File), 51KB - jmir_v21i1e10421_app1.pdf]

Multimedia Appendix 2

The coded qualitative data.

[PDF File (Adobe PDF File), 116KB - jmir_v21i1e10421_app2.pdf]

Multimedia Appendix 3

Behavior change technique definitions from Michie et al (2013) behavior change technique taxonomy version 1.

[PDF File (Adobe PDF File), 56KB - jmir_v21i1e10421_app3.pdf]

References

1. World Health Organization. 2016. Global report on diabetes URL: http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf [accessed 2018-03-08] [WebCite Cache ID 6xlmM1Ow7]
2. IDF Diabetes Atlas, 8th ed. Brussels, Belgium: International Diabetes Federation; 2017.
3. National Diabetes Audit. NHS Digital. National diabetes audit 2012-2013 URL: https://files.digital.nhs.uk/publicationimport/pub14xxx/pub14970/nati-diab-audi-12-13-care-proc-rep.pdf and the archived link is http://www.webcitation.org/74pbHGmRe [accessed 2018-03-08] [WebCite Cache ID 74pbHGmRe]
4. Cramer JA, Benedict A, Muszbek N, Keskinaslan A, Khan ZM. The significance of compliance and persistence in the treatment of diabetes, hypertension and dyslipidaemia: a review. Int J Clin Pract 2008 Jan;62(1):76-87 [FREE Full text] [doi: 10.1111/j.1742-1241.2007.01630.x] [Medline: 17983433]
5. Farr AM, Sheehan JJ, Curkendall SM, Smith DM, Johnston SS, Kalsekar I. Retrospective analysis of long-term adherence to and persistence with DPP-4 inhibitors in US adults with type 2 diabetes mellitus. Adv Ther 2014 Dec;31(12):1287-1305 [FREE Full text] [doi: 10.1007/s12325-014-0171-3] [Medline: 25504156]
6. Krass I, Schieback P, Dhippayom T. Adherence to diabetes medication: a systematic review. Diabet Med 2015 Jun;32(6):725-737. [doi: 10.1111/dme.12651] [Medline: 25440507]
7. Hertz RP, Unger AN, Lustik MB. Adherence with pharmacotherapy for type 2 diabetes: a retrospective cohort study of adults with employer-sponsored health insurance. Clin Ther 2005 Jul;27(7):1064-1073. [doi: 10.1016/j.clinthera.2005.07.009] [Medline: 16154485]
8. Farmer AJ, Rodgers LR, Lonergan M, Shields B, Weedon MN, Donnelly L, et al. Adherence to oral glucose-lowering therapies and associations with 1-year HbA1c: a retrospective cohort analysis in a large primary care database. Diabetes Care 2016 Feb;39(2):258-263 [FREE Full text] [doi: 10.2337/dc15-1194] [Medline: 26681714]
9. Pladevall M, Williams LK, Potts LA, Divine G, Xi H, Lafata JE. Clinical outcomes and adherence to medications measured by claims data in patients with diabetes. Diabetes Care 2004 Dec;27(12):2800-2805 [FREE Full text] [Medline: 15562188]
10. Egede LE, Gebregziabher M, Dismuke CE, Lynch CP, Axon RN, Zhao Y, et al. Medication nonadherence in diabetes: longitudinal effects on costs and potential cost savings from improvement. Diabetes Care 2012 Dec;35(12):2533-2539 [FREE Full text] [doi: 10.2337/dc12-0572] [Medline: 22912429]
11. Kanavos P, van den Aardweg S, Schurer W. LSE Health. London: London School of Economics; 2012 Jan. Diabetes expenditure, burden of disease and management in 5 EU countries URL: http://www.lse.ac.uk/business-and-consultancy/consulting/assets/documents/diabetes-expenditure-burden-of-disease-and-management-in-5-eu-countries.pdf [accessed 2018-11-16] [WebCite Cache ID 73xvYayyp]
12. Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RC, et al. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review. Ann Intern Med 2012 Dec 4;157(11):785-795. [doi: 10.7326/0003-4819-157-11-201212040-00538] [Medline: 22964778]
13. Nieuwlaat R, Wilczynski N, Navarro T, Hobson N, Jeffery R, Keepanasseril A, et al. Interventions for enhancing medication adherence. Cochrane Database Syst Rev 2014;11:CD000011. [doi: 10.1002/14651858.CD000011.pub4] [Medline: 25412402]
14. Doggrell SA. Does intervention by an allied health professional discussing adherence to medicines improve this adherence in Type 2 diabetes? Diabet Med 2010 Dec;27(12):1341-1349. [doi: 10.1111/j.1464-5491.2010.03137.x] [Medline: 21059085]
15. Lindenmeyer A, Hearnshaw H, Vermeire E, Van Royen P, Wens J, Biot Y. Interventions to improve adherence to medication in people with type 2 diabetes mellitus: a review of the literature on the role of pharmacists. J Clin Pharm Ther 2006 Oct;31(5):409-419. [doi: 10.1111/j.1365-2710.2006.00759.x] [Medline: 16958818]
16. Omran D, Guirguis LM, Simpson SH. Systematic review of pharmacist interventions to improve adherence to oral antidiabetic medications in people with type 2 diabetes. Can J Diabetes 2012 Oct;36(5):292-299. [doi: 10.1016/j.jcjd.2012.07.002]

17. Sapkota S, Brien J, Greenfield J, Aslani P. A systematic review of interventions addressing adherence to anti-diabetic medications in patients with type 2 diabetes--impact on adherence. PLoS One 2015 Feb 24;10(2):e0118296 [FREE Full text] [doi: 10.1371/journal.pone.0118296] [Medline: 25710465]
18. Wens J, Vermeire E, Hearnshaw H, Lindenmeyer A, Biot Y, Van Royen P. Educational interventions aiming at improving adherence to treatment recommendations in type 2 diabetes: a sub-analysis of a systematic review of randomised controlled trials. Diabetes Res Clin Pract 2008 Mar;79(3):377-388. [doi: 10.1016/j.diabres.2007.06.006] [Medline: 17643546]
19. International Telecommunications Union. 2014. ICT facts and figures URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2014-e.pdf [accessed 2018-11-13] [WebCite Cache ID 73tfHRpYt]
20. Office of National Statistics. 2014. List of tables: Appendix A URL: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/compendium/familyspending/2015/listoftablesappendixa [accessed 2018-03-08] [WebCite Cache ID 6xlmru9IO]
21. Office of National Statistics. 2017. Internet access - households and individuals URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017 [accessed 2018-03-08] [WebCite Cache ID 6xlmvP7GP]
22. Pew Research Centre. 2017. Mobile fact sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-03-08] [WebCite Cache ID 6xln63MWb]
23. Hurling R, Fairley BW, Dias MB. Internet-based exercise intervention systems: are more interactive designs better? Psychol Health 2006 Dec;21(6):757-772. [doi: 10.1080/14768320600603257]
24. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007 Apr;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
25. Moore MJ, Soderquist J, Werch C. Feasibility and efficacy of a binge drinking prevention intervention for college students delivered via the Internet versus postal mail. J Am Coll Health 2005;54(1):38-44. [doi: 10.3200/JACH.54.1.38-44] [Medline: 16050327]
26. Farmer AJ, McSharry J, Rowbotham S, McGowan L, Ricci-Cabello I, French DP. Effects of interventions promoting monitoring of medication use and brief messaging on medication adherence for people with Type 2 diabetes: a systematic review of randomized trials. Diabet Med 2015 Oct 15;33(5):565-579. [doi: 10.1111/dme.12987] [Medline: 26470750]
27. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
28. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Jan;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
29. The Center for Research in Implementation Science and Prevention. 2013. Text messaging in healthcare research toolkit URL: http://www.ucdenver.edu/academics/colleges/medicalschool/programs/crisp/training/toolkits/textingtoolkit/Documents/Text%20Messaging%20in%20Healthcare%20Research%20Toolkit%202.pdf [accessed 2018-03-08] [WebCite Cache ID 6xlniD5Ax]
30. Kelly SE, Moher D, Clifford TJ. Defining rapid reviews: a modified Delphi consensus approach. Int J Technol Assess Health Care 2016 Jan;32(4):265-275. [doi: 10.1017/S0266462316000489] [Medline: 27724987]
31. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Appl Psychology 2008 Oct 1;57(4):660-680. [doi: 10.1111/j.1464-0597.2008.00341.x]
32. Bartholomew L, Markham C, Ruiter R, Kok G, Parcel G. Chapter 1. Overview of intervention mapping. In: Planning health promotions: an intervention mapping approach. Hoboken, NJ: John Wiley & Sons; 2016:21-22.
33. Broekmans S, Dobbels F, Milisen K, Morlion B, Vanderschueren S. Medication adherence in patients with chronic non-malignant pain: is there a problem? Eur J Pain 2009 Feb;13(2):115-123. [doi: 10.1016/j.ejpain.2008.02.010] [Medline: 18467138]
34. Fogarty L, Roter D, Larson S, Burke J, Gillespie J, Levy R. Patient adherence to HIV medication regimens: a review of published and abstract reports. Patient Educ Couns 2002 Feb;46(2):93-108. [Medline: 11867239]
35. Ng CJ, Lai PS, Lee YK, Azmi SA, Teo CH. Barriers and facilitators to starting insulin in patients with type 2 diabetes: a systematic review. Int J Clin Pract 2015 Oct;69(10):1050-1070. [doi: 10.1111/ijcp.12691] [Medline: 26147376]
36. Conn VS, Hafdahl AR, Cooper PS, Ruppar TM, Mehr DR, Russell CL. Interventions to improve medication adherence among older adults: meta-analysis of adherence outcomes among randomized controlled trials. Gerontologist 2009 Aug;49(4):447-462. [doi: 10.1093/geront/gnp037] [Medline: 19460887]
37. Conn VS, Ruppar TM, Enriquez M, Cooper PS, Chan KC. Healthcare provider targeted interventions to improve medication adherence: systematic review and meta-analysis. Int J Clin Pract 2015 Aug;69(8):889-899 [FREE Full text] [doi: 10.1111/ijcp.12632] [Medline: 25728214]

38. Conn VS, Ruppar TM, Chase JD, Enriquez M, Cooper PS. Interventions to improve medication adherence in hypertensive patients: systematic review and meta-analysis. Curr Hypertens Rep 2015 Dec;17(12):94 [FREE Full text] [doi: 10.1007/s11906-015-0606-5] [Medline: 26560139]
39. Conn VS, Ruppar TM, Enriquez M, Cooper P. Medication adherence interventions that target subjects with adherence problems: systematic review and meta-analysis. Res Social Adm Pharm 2016 Mar;12(2):218-246 [FREE Full text] [doi: 10.1016/j.sapharm.2015.06.001] [Medline: 26164400]
40. Conn VS, Ruppar TM. Medication adherence outcomes of 771 intervention trials: systematic review and meta-analysis. Prev Med 2017 Jun;99:269-276. [doi: 10.1016/j.ypmed.2017.03.008] [Medline: 28315760]
41. Cutrona SL, Choudhry NK, Stedman M, Servi A, Liberman JN, Brennan T, et al. Physician effectiveness in interventions to improve cardiovascular medication adherence: a systematic review. J Gen Intern Med 2010 Oct;25(10):1090-1096 [FREE Full text] [doi: 10.1007/s11606-010-1387-9] [Medline: 20464522]
42. Devine EC, Reifschneider E. A meta-analysis of the effects of psychoeducational care in adults with hypertension. Nurs Res 1995 Aug;44(4):237-245. [Medline: 7624235]
43. Dew MA, DiMartini AF, De Vito Dabbs A, Myaskovsky L, Steel J, Unruh M, et al. Rates and risk factors for nonadherence to the medical regimen after adult solid organ transplantation. Transplantation 2007 Apr 15;83(7):858-873. [doi: 10.1097/01.tp.0000258599.65257.a6] [Medline: 17460556]
44. Holmes EA, Hughes DA, Morrison VL. Predicting adherence to medications using health psychology theories: a systematic review of 20 years of empirical research. Value Health 2014 Dec;17(8):863-876 [FREE Full text] [doi: 10.1016/j.jval.2014.08.2671] [Medline: 25498782]
45. Kahwati L, Viswanathan M, Golin CE, Kane H, Lewis M, Jacobs S. Identifying configurations of behavior change techniques in effective medication adherence interventions: a qualitative comparative analysis. Syst Rev 2016 May 04;5(1):83 [FREE Full text] [doi: 10.1186/s13643-016-0255-z] [Medline: 27209092]
46. Ruppar TM, Delgado JM, Temple J. Medication adherence interventions for heart failure patients: a meta-analysis. Eur J Cardiovasc Nurs 2015 Oct;14(5):395-404. [doi: 10.1177/1474515115571213] [Medline: 25669661]
47. Schedlbauer A, Davies P, Fahey T. Interventions to improve adherence to lipid lowering medication. Cochrane Database Syst Rev 2010 Mar 17(3):CD004371. [doi: 10.1002/14651858.CD004371.pub3] [Medline: 20238331]
48. Simoni JM, Pearson CR, Pantalone DW, Marks G, Crepaz N. Efficacy of interventions in improving highly active antiretroviral therapy adherence and HIV-1 RNA viral load. A meta-analytic review of randomized controlled trials. J Acquir Immune Defic Syndr 2006 Dec 1;43(Suppl 1):S23-S35 [FREE Full text] [doi: 10.1097/01.qai.0000248342.05438.52] [Medline: 17133201]
49. Takiya LN, Peterson AM, Finley RS. Meta-analysis of interventions for medication adherence to antihypertensives. Ann Pharmacother 2004 Oct;38(10):1617-1624. [doi: 10.1345/aph.1D268] [Medline: 15304624]
50. Teeter BS, Kavookjian J. Telephone-based motivational interviewing for medication adherence: a systematic review. Transl Behav Med 2014 Dec;4(4):372-381 [FREE Full text] [doi: 10.1007/s13142-014-0270-3] [Medline: 25584086]
51. Thorneloe RJ, Bundy C, Griffiths CE, Ashcroft DM, Cordingley L. Adherence to medication in patients with psoriasis: a systematic literature review. Br J Dermatol 2013 Jan 01;168(1):20-31. [doi: 10.1111/bjd.12039] [Medline: 22963128]
52. Xu A, Chomutare T, Iyengar S. Persuasive attributes of medication adherence interventions for older adults: a systematic review. Technol Health Care 2014 Jan 01;22(2):189-198. [doi: 10.3233/THC-140812] [Medline: 24837055]
53. Vignon Zomahoun HT, de Bruin M, Guillaumie L, Moisan J, Grégoire J, Pérez N, et al. Effectiveness and content analysis of interventions to enhance oral antidiabetic drug adherence in adults with type 2 diabetes: systematic review and meta-analysis. Value Health 2015 Jun 01;18(4):530-540 [FREE Full text] [doi: 10.1016/j.jval.2015.02.017] [Medline: 26091607]
54. Brundisini F, Vanstone M, Hulan D, DeJean D, Giacomini M. Type 2 diabetes patients' and providers' differing perspectives on medication nonadherence: a qualitative meta-synthesis. BMC Health Serv Res 2015 Nov 23;15:516 [FREE Full text] [doi: 10.1186/s12913-015-1174-8] [Medline: 26596271]
55. Kumar K, Greenfield S, Raza K, Gill P, Stack R. Understanding adherence-related beliefs about medicine amongst patients of South Asian origin with diabetes and cardiovascular disease patients: a qualitative synthesis. BMC Endocr Disord 2016 May 26;16(1):24 [FREE Full text] [doi: 10.1186/s12902-016-0103-0] [Medline: 27230479]
56. McSharry J, McGowan L, Farmer AJ, French DP. Perceptions and experiences of taking oral medications for the treatment of Type 2 diabetes mellitus: a systematic review and meta-synthesis of qualitative studies. Diabet Med 2016 Dec;33(10):1330-1338. [doi: 10.1111/dme.13152] [Medline: 27150899]
57. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York: University of York; Jan 2009.
58. Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. Implement Sci 2010 Jul 19;5:56 [FREE Full text] [doi: 10.1186/1748-5908-5-56] [Medline: 20642853]
59. Harker J, Kleijnen J. What is a rapid review? A methodological exploration of rapid reviews in Health Technology Assessments. Int J Evid Based Healthc 2012 Dec 01;10(4):397-410. [doi: 10.1111/j.1744-1609.2012.00290.x] [Medline: 23173665]

60. Khangura S, Polisena J, Clifford TJ, Farrah K, Kamel C. Rapid review: an emerging approach to evidence synthesis in health technology assessment. Int J Technol Assess Health Care 2014 Jan;30(1):20-27. [doi: 10.1017/S0266462313000664] [Medline: 24451157]
61. Bero L, Busuttil G, Farquhar C, Koehlmoos TP, Moher D, Nylenna M, et al. Measuring the performance of the Cochrane library. Cochrane Database Syst Rev 2012 Nov 14;12:ED000048. [doi: 10.1002/14651858.ED000048] [Medline: 23235688]
62. Schünemann HJ, Moja L. Reviews: Rapid! Rapid! Rapid! …and systematic. Syst Rev 2015 Jan 14;4:4 [FREE Full text] [doi: 10.1186/2046-4053-4-4] [Medline: 25589399]
63. Pai M, McCulloch M, Gorman JD, Pai N, Enanoria W, Kennedy G, et al. Systematic reviews and meta-analyses: an illustrated, step-by-step guide. Natl Med J India 2004;17(2):86-95. [Medline: 15141602]
64. Michie S, Carey RN, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. From theory-inspired to theory-based interventions: a protocol for developing and testing a methodology for linking behaviour change techniques to theoretical mechanisms of action. Ann Behav Med 2016 Jul 11;18(4):530-540. [doi: 10.1007/s12160-016-9816-6] [Medline: 27401001]

Abbreviations

BCT: behavior change technique
BCTT: behavior change technique taxonomy
BS: behavioral strategy
ES: effect size
MA: medication adherence
MeSH: Medical Subject Heading
mHealth: mobile health
NIHR: National Institute for Health Research
RCT: randomized controlled trial
SMD: standard mean difference
SMS: short message service
TC: theoretical construct

Edited by G Eysenbach; submitted 16.03.18; peer-reviewed by L Avery, C Osborn, S Skovlund; comments to author 31.07.18; revised version received 25.09.18; accepted 03.10.18; published 09.01.19.

Please cite as:
Long H, Bartlett YK, Farmer AJ, French DP
Identifying Brief Message Content for Interventions Delivered via Mobile Devices to Improve Medication Adherence in People With Type 2 Diabetes Mellitus: A Rapid Systematic Review
J Med Internet Res 2019;21(1):e10421
URL: https://www.jmir.org/2019/1/e10421/
doi:10.2196/10421
PMID:30626562

Original Paper

Evaluating a Web-Based Social Anxiety Intervention Among Community Users: Analysis of Real-World Data

Hugh Cameron McCall[1], BA; Fjola Dogg Helgadottir[2,3,4], PhD; Ross G Menzies[5], PhD; Heather D Hadjistavropoulos[1], PhD; Frances S Chen[6], PhD

[1]Department of Psychology, University of Regina, Regina, SK, Canada

[2]AI-Therapy, North Vancouver, BC, Canada

[3]Dr. Fjóla & Kompaní, Reykjavík, Iceland

[4]Hugræna Atferlisstöðin í Reykjavík, Reykjavík, Iceland

[5]Graduate School of Health, University of Technology Sydney, Sydney, Australia

[6]Department of Psychology, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Fjola Dogg Helgadottir, PhD
AI-Therapy
5515 Stellar Place
North Vancouver, BC, V7R 4N3
Canada
Phone: 1 604 652 5440
Email: fjola@ai-therapy.com

Abstract

Background: Social anxiety is both harmful and prevalent. It also currently remains among the most undertreated major mental disorders, due, in part, to socially anxious individuals' concerns about the stigma and expense of seeking help. The privacy and affordability of computer-aided psychotherapy interventions may render them particularly helpful in addressing these concerns, and they are also highly scalable, but most tend to be only somewhat effective without therapist support. However, a recent evaluation of a new self-guided, 7-module internet-delivered cognitive behavioral therapy intervention called Overcome Social Anxiety found that it was highly effective.

Objective: The initial evaluation of Overcome Social Anxiety revealed that it led to significant reductions in symptom severity among university undergraduates. The aim of this study was to extend the results of the initial study and investigate their generalizability by directly evaluating the intervention's effectiveness among a general community sample.

Methods: While signing up for Overcome Social Anxiety, users consented to the usage of their anonymized outcome data for research purposes. Before and after completing the intervention, users completed the Fear of Negative Evaluation Scale (FNE), which we employed as the primary outcome measure. Secondary outcome measures included the Depression Anxiety Stress Scales (DASS) and 2 bespoke questionnaires measuring socially anxious thoughts (Thoughts Questionnaire) and avoidance behaviors (Avoidance Questionnaire).

Results: Participants who completed the intervention (102/369, 27.7%) experienced significant reductions in the severity of their symptoms on all measures employed, including FNE (P<.001; Cohen d=1.76), the depression subscale of DASS (P<.001; Cohen d=0.70), the anxiety subscale of DASS (P<.001; Cohen d=0.74), the stress subscale of DASS (P<.001; Cohen d=0.80), the Thoughts Questionnaire (P<.001; Cohen d=1.46), and the Avoidance Questionnaire (P<.001; Cohen d=1.42).

Conclusions: Our results provide further evidence that Overcome Social Anxiety reduces the severity of social anxiety symptoms among those who complete it and suggest that its effectiveness extends to the general community. The completion rate is the highest documented for a fully automated intervention for anxiety, depression, or low mood in a real community sample. In addition, our results indicate that Overcome Social Anxiety reduces the severity of symptoms of depression, physiological symptoms of anxiety, and stress in addition to symptoms of social anxiety.

(J Med Internet Res 2019;21(1):e11566) doi:10.2196/11566

KEYWORDS

social anxiety; internet; cognitive behavioral therapy; psychotherapy; mental health

Introduction

Background

Social anxiety disorder has a high lifetime prevalence of approximately 13% [1]. It causes considerable distress and functional impairment, even at a subclinical level of severity [2], and impacts both the private and professional lives of those affected by it [3,4]. It is persistent in the absence of treatment [5] and is related to other mental disorders such as mood and substance disorders [6]. Yet, social anxiety remains one of the most undertreated of all major mental disorders today [7]. Importantly, its relatively low treatment rate cannot be attributed to any lack of empirically supported treatment methods; research has shown that both psychotherapeutic treatments (eg, cognitive behavioral therapy, CBT) and pharmaceutical treatments (eg, selective serotonin reuptake inhibitors) for social anxiety are effective [8,9]. Rather, financial constraints and concerns about being judged or stigmatized for seeking help, among other issues, represent major barriers to treatment for socially anxious individuals [7].

Computerized cognitive behavioral therapy (CCBT), a promising and increasingly popular treatment for anxiety and depression [10], may be particularly useful in surmounting social anxiety's unique barriers to treatment, as it is more affordable and can be accessed more privately than traditional psychotherapy. Most computer-aided psychotherapy (CP) interventions (a category of interventions including CCBT) involve some therapist support, and the effectiveness of CP interventions is related to the amount of therapist support their users receive [11,12]. Accordingly, self-guided CCBT interventions—those designed to operate independently, without the necessity of therapist support—tend to be less effective than those involving therapist support [13].

Overcome Social Anxiety

A recent study suggests that a self-guided internet-delivered cognitive behavioral therapy (ICBT; CCBT delivered via the internet) intervention called Overcome Social Anxiety may represent a notable exception to the tendency for self-guided CCBT interventions to be less effective than those involving therapist support [14]. The study—a randomized controlled trial among undergraduate university students, which compared Overcome Social Anxiety with a wait-list control condition—revealed a between-groups effect size (Cohen *d*=0.97) similar to the average between-groups effect size of 19 trials of computer-aided interventions *with* therapist support found in a review (Cohen *d*=0.96) [11].

Overcome Social Anxiety arose from a program of research exploring the common limitations of other CCBT interventions [15] and was designed to address 5 such limitations in particular. First, where many other interventions do not adequately individualize treatment to individual participants' needs, Overcome Social Anxiety employs a series of questionnaires to tailor each user's treatment package to address their unique symptoms and the contexts in which those symptoms typically occur. Second, corrective feedback on important aspects of the CBT process is often lacking from CCBT interventions; Overcome Social Anxiety mitigates the need for corrective feedback by providing example responses to help ensure that users fully understand what they are required to do at each stage of the treatment process (eg, challenging maladaptive thoughts and designing behavioral experiments). Third, where some interventions do not adequately address low adherence rates [15], which remain common among CP interventions today [16], Overcome Social Anxiety employs 2 mechanisms to encourage users to make steady progress: (1) it hinders procrastination by limiting users to a 6-month window to complete the intervention and (2) it mitigates forgetfulness by sending users automated reminders to continue their work on the program following periods of inactivity. Fourth, although research shows that therapist-client interaction is an important aspect of successful CBT [17], it is by definition lacking from stand-alone CCBT interventions. Overcome Social Anxiety addresses this challenge by employing voice recordings of 2 clinical psychologists to guide users through the treatment process, more closely mirroring a traditional course of human-delivered therapy. A recent study found that the *patient-program alliance*, similar to the therapeutic alliance, is associated with greater adherence and more favorable clinical outcomes in CCBT [18], indicating that such efforts at improving how users relate to interventions themselves are unlikely to be misguided. Finally, many other interventions have failed to provide users with a sufficient dose of treatment to effect lasting positive change [15], despite research attesting to the importance of an appropriate dose of treatment to CBT's success [19]. Overcome Social Anxiety was designed to deliver a more robust treatment package than many other programs and includes all established elements of modern CBT.

Overcome Social Anxiety comprises an assessment battery and 7 core modules. The intervention begins with a *prequestionnaires* module, which is designed both to take a pretreatment measure of users' symptom severity and to individualize the treatment to each user. Module 1, *Thinking exercises*, introduces users to the program and their virtual psychologists, informs users of common cognitive errors, and explains the relationship between cognitions, behaviors, and emotions. Module 2, *Challenging your thinking*, presents users with personally relevant anxious thoughts (based on prequestionnaire responses) and asks users to challenge those thoughts through writing exercises. In module 3, *Creating your model*, users select symptoms and anxiety-inducing situations and cognitions, which the intervention then uses to individualize the treatment to users' unique experiences of social anxiety. This model is then applied to module 4, *Behavioral experiments*, wherein users are guided through a series of behavioral experiments to target safety behaviors and avoidance. Module 5, *Challenge your thinking further*, continues to help users adjust their negative beliefs, with a particular emphasis on anger. Module 6, *Self-processing*, targets biased attentional processes through skills-based attention training [20] and rescripting of faulty and negative imagery [21]. Module 7, *Relapse prevention*, reviews the material covered in the first 6 core modules and provides users with psychoeducation to help them maintain treatment gains. Finally, users complete the *postquestionnaires* module, which the program uses to create histograms to show users the difference between their pre- and posttreatment symptom severity. The program then provides users with an

individualized PDF containing all program materials, which users can employ to help maintain treatment gains into the future.

Each module was designed to achieve a particular clinical goal and not necessarily to be completed in a single sitting. Indeed, some require substantially more time and effort than others. For example, when users are taught how to change their thinking in module 2, they are encouraged to target 1 thought they are working on changing per day. Please see the initial trial [14] for more detailed information about the content of the program and Multimedia Appendix 1 for a screenshot.

The results of the initial trial indicate that the 5 design features discussed above may be collectively very useful, potentially contributing enough to bridge the effectiveness gap between self-guided and therapist-assisted interventions. Overcome Social Anxiety—if it is indeed as effective as the initial trial suggests—may have important implications not only for reducing the severity of symptoms and increasing the well-being of people who struggle with social anxiety but also for the development of future interventions.

Although the initial study was a randomized controlled trial with high internal validity, it did not explicitly investigate the generalizability of its findings to the general community of individuals with social anxiety. The purpose of this study was to attempt to replicate the findings of the initial trial with high external validity by employing Overcome Social Anxiety's general user base as its sample. Our rationale for conducting this study was that—to the extent that the intervention was found to be similarly effective among its general user base as it was among the initial trial's student sample—consumers, mental health professionals, researchers, and developers of future ICBT interventions would be able to place more confidence in the intervention's effectiveness.

Hypothesis

We hypothesized that among those who completed all modules of the intervention (including its prequestionnaires module, its 7 core modules, and its postquestionnaires module), referred to throughout this study as *completers*, posttreatment scores on the Fear of Negative Evaluation Scale (FNE) [22] would be significantly lower than pretreatment scores. The results from the initial trial [14], which also employed FNE and found a large pretreatment-to-posttreatment effect size for that measure (Cohen *d*=0.82), suggested that this difference could be large.

Methods

Participants

In this study, we retrospectively analyzed data from past users of Overcome Social Anxiety. These data were automatically collected from users between August 2012 and April 2018. Thus, no new data collection was necessary. Our sample (n=369) consisted of all former, paying users of Overcome Social Anxiety. We excluded (1) those whose usage of the program was ongoing as of April 2018, (2) the university undergraduates who participated in the initial trial [14], and (3) all other users who were given free access to the intervention (eg, through the private practices of its creators).

A power analysis suggested that, for within-subjects comparisons of pre- and posttreatment FNE scores, a sample of 40 completers would have been required to achieve a power level of 0.99 at the $P<.01$ level of significance, two-tailed, assuming the effect size of Cohen $d=0.82$ found in the initial trial [14]. Even given a conservative estimate of an effect half this size (ie, Cohen $d=0.41$) for this study's population, we would have achieved a power level of 0.93 at the $P<.01$ level of significance, two-tailed, with the 102 users who completed the intervention.

All past and present users of Overcome Social Anxiety consented to the collection, anonymization, and later analysis of their data for research purposes during registration. The protocol for this study was approved by the University of British Columbia's Behavioural Research Ethics Board (Human Ethics Application ID: H16-00319).

Outcome Measures

Overcome Social Anxiety begins with a prequestionnaires module, which is included both to individualize the course of treatment to each user's needs and to measure each user's pretreatment symptom severity. The intervention concludes with a postquestionnaires module, which allows users to quantify changes in the severity of their symptoms. Both these modules contain 4 measures. First, they contain the FNE scale [22], a well-validated measure of social anxiety symptoms [22,23]. The brief FNE scale, whose scores have been found to share a Pearson correlation coefficient of .96 with the original FNE ($r=.96$) [24], correlates with other measures of social anxiety and yields significantly higher scores among those diagnosed with social anxiety disorder than those without [25]. The FNE comprises 30 statements (eg, "I worry that others will think I am not worthwhile") that respondents mark as true or false and yields a total score between 0 and 30. Scores of 7 (1 SD below the mean for a large student sample) and 8 (lower quartile) have been recommended as cut-off scores to indicate low social anxiety, whereas scores of 22 (1 SD above the mean) and 20 (upper quartile) have been recommended as cut-off scores to indicate high social anxiety [23].

Second, both these modules include the Depression Anxiety Stress Scales (DASS) [26], a 42-item questionnaire designed to discriminate between depression (eg, "I felt sad and depressed"), anxiety (eg, "I was aware of dryness of my mouth"), and stress (eg, "I was in a state of nervous tension"), despite their shared symptoms. Research has attested to its reliability and validity [27,28].

Finally, these modules include 2 bespoke questionnaires, which ask users about the frequency with which they experience 37 socially anxious thoughts (eg, "I can't speak to authority figures") and avoid 23 anxiety-provoking situations (eg, "Making small talk with strangers/colleagues"). The items on these 2 questionnaires—titled the Thoughts Questionnaire and the Avoidance Questionnaire, respectively—were retrieved from a file audit of decades of clinical psychology practice with individuals diagnosed with social anxiety. Both questionnaires are scored on 5-point Likert scales and are intended to capture the patterns of thinking and behavior characteristic of real experience with social anxiety symptoms. Cronbach alphas for

the Thoughts Questionnaire and the Avoidance Questionnaire were .94 and .89, respectively. In our sample, participants' scores on FNE were related to their scores on both bespoke questionnaires, with Pearson correlation coefficients of .58 (n=369; *P*<.001) for the Thoughts Questionnaire and .36 (n=369; *P*<.001) for the Avoidance Questionnaire.

Procedure

Users from around the world found Overcome Social Anxiety independently and chose to sign up for US $149.99. The intervention was advertised through Google Adwords between August 2012 and August 2014, has received media coverage, and has seen mention in blog posts, which likely helped users to discover it. During registration, users consented to the use of their data for research. After signing up, users first completed the prequestionnaires module, consisting of FNE, DASS, the Thoughts Questionnaire, and the Avoidance Questionnaire. They then began working through the 7 core modules of the program (see Figure 1 for content outline and the initial trial [14] for details). On completing all the program's core modules, users responded to FNE, DASS, the Thoughts Questionnaire, and the Avoidance Questionnaire again during the postquestionnaires module. At the end of this module, they were also asked to leave feedback for the creators of the intervention. Users were given a limit of 6 months from the date of their registration to complete the intervention. They were sent automated emails reminding them to continue using the program after 3, 7, 10, 14, 21, and 28 days of inactivity. On completing each module, users were also sent automated emails summarizing that module's contents. As this study involved only the retrospective analysis of data from past users of Overcome Social Anxiety, there was no contact with participants throughout the course of the study.

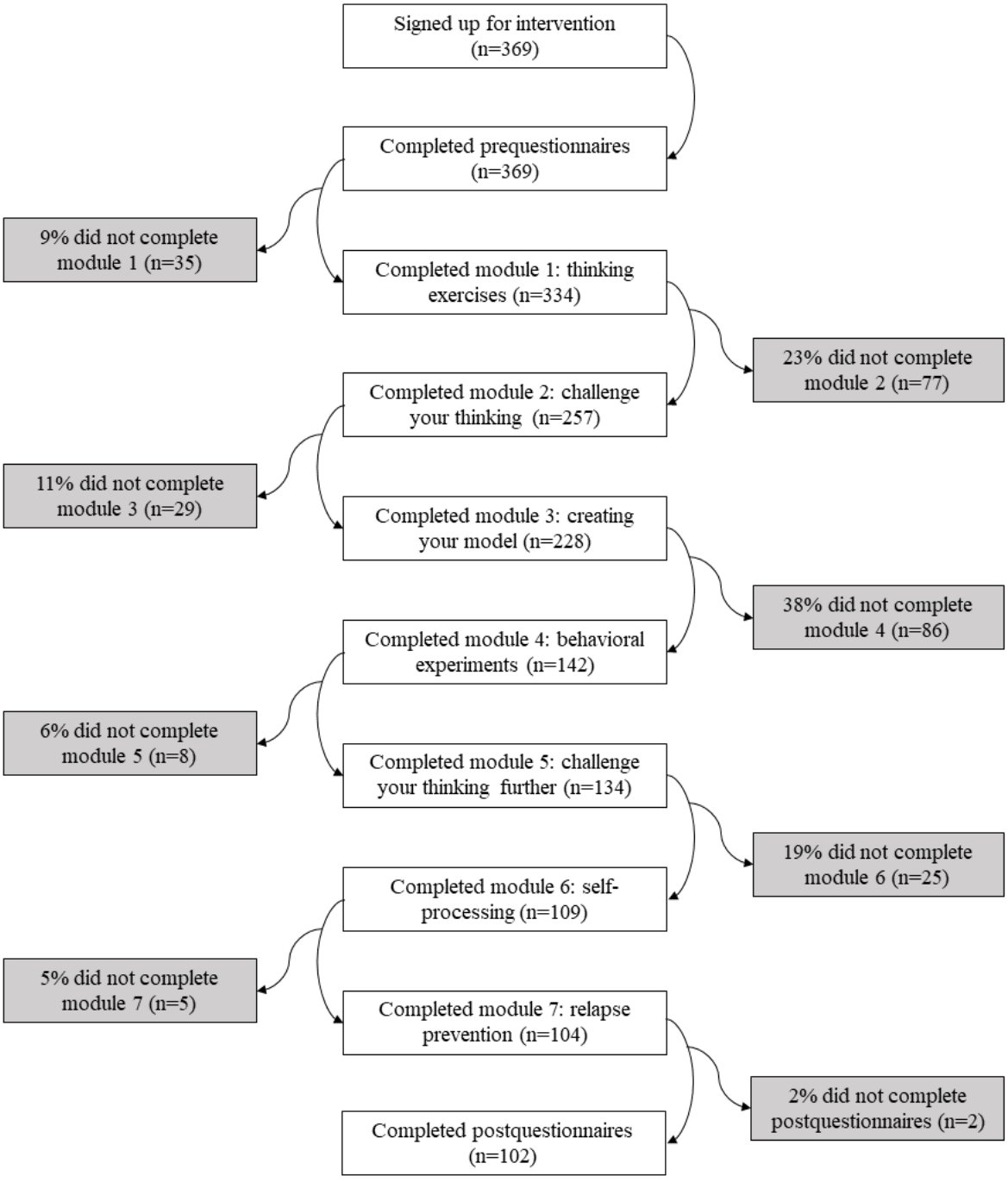
Analyses

This study's primary dependent variable was pretreatment-to-posttreatment change in the severity of social anxiety symptoms, as measured by FNE. We selected FNE as a primary measure because it is a well-established and validated measure of social anxiety symptoms and because our other measure of anxiety, the anxiety subscale of DASS, measures primarily physiological symptoms of anxiety and does not specifically measure social anxiety symptoms. However, we also analyzed changes in scores on 5 secondary measures—the 3 factors of DASS, the Thoughts Questionnaire, and the Avoidance Questionnaire. For each of these measures, within-subjects *t* tests were conducted to determine whether posttreatment scores differed from pretreatment scores among users who completed the intervention.

We tested for attrition bias by exploring differences between completers and noncompleters. Specifically, we conducted between-subjects *t* tests to check for differences between completers and noncompleters in age and pretreatment scores on all questionnaire measures. We also conducted Chi-square analyses for sex and whether users reported having previously seen a therapist for anxiety, seen a therapist for other reasons, or taken medication for anxiety. Finally, after imputing missing posttreatment data from noncompleters using a last observation carried forward approach, we conducted between-subjects *t* tests to check for pretreatment-to-posttreatment changes in symptoms on all outcome measures among all users, completers and noncompleters alike.

At the end of the postquestionnaires module, completers were asked to leave positive and negative feedback about the intervention. Employing a conventional content analysis approach [29] and QSR NVivo 11 software, we used both qualitative methods (coding participants' responses and grouping them into themes) and quantitative methods (counting comments reflecting each theme and calculating descriptive statistics) to analyze participants' responses [30]. Two researchers (KG—see Acknowledgments below—and author HM) carefully read each comment and collaboratively created a coding guide. KG used this guide to code each response and then met HM again to discuss potential inconsistencies in the application of the coding guide to the analysis of participants' comments. Finally, an expert coder (author HH) conducted a final review of the coded comments, counted comments reflecting each theme, and calculated descriptive statistics.

Figure 1. Flowchart of user progress.



Results

Program Usage

Of the 369 users who signed up for Overcome Social Anxiety between August 2012 and April 2018, 102 users (102/369, 27.7%) fully completed the intervention. The number of users who completed each module of the intervention is displayed in Figure 1. The average time taken for completion of each module, in minutes, was 14.70 for the prequestionnaires, 34.60 for module 1, 83.70 for module 2, 43.10 for module 3, 103.72 for module 4, 15.02 for module 5, 25.05 for module 6, 8.05 for module 7, and 19.10 for the postquestionnaires. Completers spent a mean of 5 hours and 34 min (SD 211 min) using the intervention, whereas noncompleters spent a mean of 2 hours and 11 min (SD 86 min). However, these data represent only the amount of time users spent logged in to the intervention. Important components of CBT, including homework exercises

such as exposure activities, occur between sessions, and the overall amount of time spent is likely to be considerably higher. On average, completers logged in 25.06 times (SD 17.25) over a period of 138.07 days (SD 126.07), whereas noncompleters logged in 9.93 times (SD 8.14) over a period of 58.02 days (SD 83.27).

User Characteristics

Out of the 102 completers, 39 (38.2%) identified as female, 58 (56.9%) identified as male, and 5 (4.9%) did not report their sex. The mean age of completers was 35.47 (SD 13.64). When asked about their clinical history, many completers reported having previously seen a therapist for anxiety (45/94, 48%), seen a therapist for another reason (43/95, 45%), and taken medication for anxiety (36/96, 38%). These data, in addition to those of noncompleters, are displayed in Table 1. The most common countries of residence users reported during registration were the United States (105/369, 28.5%), the United Kingdom (56/369, 15.2%), Australia (55/369, 15.0%), Iceland (30/369, 8.1%), and Canada (28/369, 7.6%). The remaining users were spread across 25 other countries around the world (60/369, 16.3%) or did not report their countries of residence (35/369, 9.5%).

Pretreatment Questionnaire Scores

The mean pretreatment FNE score among completers was 25.91 (SD 3.99), indicating very high levels of social anxiety. This score approached the FNE's maximum score of 30 and exceeded cut-off scores defining high anxiety (20 and 22) by a considerable margin [23]. For DASS, completers had pretreatment scores of 14.07 (SD 10.07) on the depression subscale, 9.93 (SD 6.84) on the anxiety subscale, and 16.28 (SD 8.05) on the stress subscale. Finally, completers' mean pretreatment scores on the Thoughts Questionnaire and Avoidance Questionnaire were 79.72 (SD 23.4) and 51.37 (SD 14.85), respectively.

Comparison of Completers and Noncompleters

A between-subjects *t* test revealed that completers had lower scores than noncompleters on the anxiety subscale of DASS (equal variances not assumed; $t_{217.66}$=2.95; *P*=.003; Cohen *d*=−0.33). Between-subjects *t* tests of DASS's depression and stress subscales, FNE, the Thoughts Questionnaire, the Avoidance Questionnaire, and age revealed no further differences between completers and noncompleters (all *P* s>.07). In addition, Chi-square analyses revealed no differences between completers and noncompleters in sex ratio or whether users reported having previously seen therapists for anxiety, seen therapists for other reasons, or taken medication for anxiety (all *P* values> .15).

Effectiveness of the Intervention

Completers experienced significant pretreatment-to-posttreatment reductions in symptom severity on all measures employed: FNE (t_{101}=13.61; *P*<.001; Cohen *d*=1.76), the depression subscale of DASS (t_{101}=7.42; *P*<.001; Cohen *d*=0.70), the anxiety subscale of DASS (t_{101}=8.24; *P*<.001; Cohen *d*=0.74), the stress subscale of DASS (t_{101}=9.57; *P*<.001; Cohen *d*=0.80), the Thoughts Questionnaire (t_{101}=16.47; *P*<.001; Cohen *d*=1.46), and the Avoidance Questionnaire (t_{101}=15.40; *P*<.001; Cohen *d*=1.42). Demographic characteristics and pretreatment questionnaire scores for all users, in addition to posttreatment questionnaire scores and symptom change analyses for completers, are summarized in Table 1.

For each outcome measure, some completers reported a worsening in symptom severity from pre- to posttreatment. These changes were all less than 1 SD (ie, 1 pretreatment SD among completers for each measure) in magnitude, except for those of 3 (3/102, 3.0%) participants for FNE, 4 (4/102, 4.0%) for the depression subscale of DASS, 1 (1/102, 1.0%) for the anxiety subscale of DASS, and 1 (1/102, 1.0%) for the stress subscale of DASS.

As discussed above in outcome measures, FNE threshold scores of 20 and 22 have been recommended to distinguish between high anxiety and moderate or low anxiety individuals [23]. According to these thresholds, 96 (96/102, 94.1%) or 90 (90/102, 88.2%) completers reported high anxiety before beginning the clinical content of the intervention, whereas 33 (33/102, 32.4%) or 28 (28/102, 27.5%) reported high anxiety after completing it. One completer began the intervention below 1 of the thresholds (with a score of 20) and ended it above that threshold (with a score of 24). All other participants either remained in the same category of anxiety or experienced a reduction in their FNE score, which moved them past a threshold and into a lower category of anxiety.

Table 1. User characteristics and questionnaire scores.

Characteristic	Completers (n=102)	Noncompleters (n=267)	Total (N=369)
User characteristics[a]			
Female, n (%)	39 (40.2)	98 (38.4)	137 (39.0)
Age, mean (SD)[b]	35.47 (13.64)	33.88 (11.93)	34.31 (12.42)
Seen therapist for anxiety, n (%)	45 (47.9)	129 (52.0)	174 (50.9)
Seen therapist for other reasons, n (%)	43 (45.3)	96 (38.9)	139 (40.6)
Taken medication for anxiety, n (%)	36 (37.5)	120 (48.6)	156 (45.5)
Fear of negative evaluation			
Pretreatment, mean (SD)	25.91 (3.99)	25.53 (5.05)	25.63 (4.78)
Posttreatment, mean (SD)	15.06 (8.32)	—	—
Change[c]			
t test (*df*)	13.61 (101)	—	—
P value	<.001	—	—
Cohen *d*	1.76	—	—
DASS[d] **(depression subscale)**			
Pretreatment, mean (SD)	14.07 (10.07)	16.34 (11.28)	15.72 (10.99)
Posttreatment, mean (SD)	7.57 (8.60)	—	—
Change[c]			
t test (*df*)	7.42 (101)	—	—
P value	<.001	—	—
Cohen *d*	0.70	—	—
DASS (anxiety subscale)			
Pretreatment, mean (SD)	9.93 (6.84)	12.42 (8.212)	11.73 (7.93)
Posttreatment, mean (SD)	5.32 (5.64)	—	—
Change[c]			
t test (*df*)	8.24 (101)	—	—
P value	<.001	—	—
Cohen *d*	0.74	—	—
DASS (stress subscale)			
Pretreatment, mean (SD)	16.28 (8.05)	18.14 (9.40)	17.63 (9.08)
Posttreatment, mean (SD)	10.01 (7.63)	—	—
Change[c]			
t test (*df*)	9.57 (101)	—	—
P value	<.001	—	—
Cohen *d*	0.80	—	—
Thoughts Questionnaire			
Pretreatment, mean (SD)	79.72 (23.40)	80.66 (23.68)	80.40 (23.57)
Posttreatment, mean (SD)	44.93 (24.17)	—	—
Change[c]			
t test (*df*)	16.47 (101)	—	—
P value	<.001	—	—

Characteristic	Completers (n=102)	Noncompleters (n=267)	Total (N=369)
Cohen *d*	1.46	—	—
Avoidance Questionnaire			
Pretreatment, mean (SD)	51.37 (14.85)	52.35 (15.59)	52.08 (15.37)
Posttreatment, mean (SD)	29.54 (15.97)	—	—
Change[c]			
t test (*df*)	15.40 (101)	—	—
P value	<.001	—	—
Cohen *d*	1.42	—	—

[a]Some users did not respond to certain questions. The 5 rows beneath this heading display responses from 97 completers and 255 noncompleters who reported their sex, 92 completers and 246 noncompleters who reported their age, 94 completers and 248 noncompleters who reported whether or not they had previously seen a therapist for anxiety, 95 completers and 247 noncompleters who reported whether or not they had previously seen a therapist for another reason, and 96 completers and 247 noncompleters who reported whether or not they had previously taken medication for anxiety. The percentages given represent percentages of respondents, not percentages of participants overall.

[b]Age was measured by year of birth, and our statistics on users' ages represent their ages at the end of the calendar years during which they began the intervention.

[c]This row displays the results of a within-subjects *t* test comparing completers' pre- and posttreatment scores.

[d]DASS: Depression Anxiety Stress Scales.

Imputation of Missing Data

It has been suggested that where dropout rates exceed 20%—and this study's dropout rate (267/369, 72.4%) did so by far—"no adequate recommendation [for replacing missing data] can be provided" [31]. However, research has demonstrated that partial completion of ICBT interventions for anxiety and depression leads to symptom reduction [32], and noncompleters in our sample spent an average of over 2 hours using Overcome Social Anxiety, suggesting that noncompleters may have benefited from the intervention. For this reason, imputation of missing data from noncompleters using a last observation carried forward approach may be conservative. A within-subjects *t* test comparing pre- and posttreatment FNE scores for all users, assuming no pretreatment-to-posttreatment change in FNE scores for noncompleters, indicated that the program was moderately effective in reducing symptoms among all users (t_{368}=8.95; *P*<.001; Cohen *d*=0.48). Even in the hypothetical and unlikely event that noncompleters experienced an increase in social anxiety symptoms equivalent to half an SD on FNE (ie, a score increase of 2.52; Cohen *d*=0.5), our results would show a small but statistically significant reduction in FNE scores among all users (t_{368}=3.08; *P*=.002; Cohen *d*=0.18).

Furthermore, within-subjects *t* tests comparing pre- and posttreatment scores on secondary outcome measures—again, assuming no change in score for noncompleters from pre- to posttreatment—showed significant, small-to-moderate reductions in symptom severity on all secondary outcome measures: the depression subscale of DASS (t_{368}=6.31; *P*<.001; Cohen *d*=0.16), the anxiety subscale of DASS (t_{368}=6.78; *P*<.001; Cohen *d*=0.16), the stress subscale of DASS (t_{368}=7.46; *P*<.001; Cohen *d*=0.19), the Thoughts Questionnaire (t_{368}=9.63; *P*<.001; Cohen *d*=0.37), and the Avoidance Questionnaire (t_{368}=9.41; *P*<.001; Cohen *d*=0.35).

Acceptability of the Intervention

Analysis of user feedback identified 9 positive themes and 8 areas for improvement. These themes, the number of participants whose comments reflected them, and examples of these comments are displayed in Table 2. It is worth noting that feedback was only obtained from users on their completion of the program and that users were asked about both what they liked and what they did not like about the intervention. Out of 102 completers, 35 users left comments. The mean number of comments coded as positive feedback was 1.77 (SD 1.17), whereas the mean number of comments coded as areas for improvement was 0.80 (SD 0.83).

Table 2. Feedback from completers.

Feedback theme	n (%)
Positive feedback	
General praise (eg, *I loved this program.*)	26 (25.5)
Specific symptom improvement (eg, *...this program truly did help me overcome a lot of the thoughts I was having.*)	18 (17.6)
Content quality (eg, *...shows deep understanding of the problems of social anxiety.*)	12 (11.8)
Components (eg, *...I find the e book invaluable...*)	9 (8.8)
Presentation (eg, *Liked A variety of pictures, sound and interaction.*)	9 (8.8)
Convenience or accessibility (eg, *Good to be able to do things completely at your own speed.*)	6 (5.9)
Cost (eg, *Not too expensive.*)	3 (2.9)
Privacy (eg, *AI Therapy's anonymous and confidential form of treatment has been wonderful...*)	3 (2.9)
Ease of use (eg, *Easy to use. Great format. Very user friendly.*)	2 (2.0)
Areas for improvement	
Components (eg, *I found some of the content helpful, but there were some bits that I didn't find help nor did it relate to me.*)	10 (9.8)
Presentation (eg, *...I wish it was a little more visually interesting.*)	7 (6.9)
Content quality (eg, *Would be good for anxious teenagers, but somewhat too simple for adults.*)	4 (3.9)
Research considerations (eg, *...questionnaires were a bit long.*)	2 (2.0)
Technical problems (eg, *Little bit glitchy on an iPad.*)	2 (2.0)
Cost (eg, *Not much more help provided than self-help books which cost a lot less.*)	1 (1.0)
Specific lack of symptom improvement (eg, *I liked it even though i dont feel much improvement in my overall case...*)	1 (1.0)
Length (eg, *...it was a little 'longer' than I expected due to having to go out and face our fears...*)	1 (1.0)

Discussion

Principal Findings

Our primary hypothesis was that those who completed the intervention would experience a significant reduction in the severity of their social anxiety symptoms, as measured by FNE [22]. This hypothesis was clearly supported by our results. In fact, the effect size for this reduction (Cohen *d*=1.76) was larger than the mean uncontrolled pretreatment-to-posttreatment effect size of *human-delivered* CBT for social anxiety found in a meta-analysis (effect size=1.04) [33]. By comparison, although this study lacked a control condition, participants in the waitlist control condition of our initial trial experienced a mean reduction in FNE scores of 0.46 (Cohen *d*=0.14) over a similar length of time [14]. Social anxiety is widely considered to have a chronic course, and clinical and epidemiological studies have reported that it has a mean duration of 10 to 24 years [34]. Given this typical time course of social anxiety, in combination with the results for the waitlist control condition of our initial trial, we infer that the reduction in severity of social anxiety symptoms in this study is attributable primarily to the intervention rather than to spontaneous remission.

Our results also show that those who completed the intervention experienced reductions not only in the severity of social anxiety symptoms, as measured by FNE, but also in the severity of symptoms of depression, physiological symptoms of anxiety, and stress, as measured by DASS, and self-reported socially anxious thoughts and avoidance behaviors, as measured by the Thoughts Questionnaire and the Avoidance Questionnaire. Given that social anxiety is related to depression [35], general anxiety [35], and stress [36] and the 2 bespoke questionnaires measured social anxiety symptoms, these results were unsurprising. Nevertheless, they were pronounced; according to Cohen guidelines for interpreting Cohen *d* [37], all these changes were large (Cohen *d*>0.8) except for the changes on the depression and anxiety subscales of DASS, which were medium (Cohen *d*>0.5) in magnitude. We must advise caution in interpreting these results, however, as they exclude data from noncompleters and there is no control condition with which to compare them.

The completion rate (102/369, 27.7%) was high in comparison with community completion rates of other self-guided CP interventions. Data from community users show lower completion rates than data from trials [16], and self-guided CP interventions have lower completion rates than therapist-assisted CP interventions [38]; it is, therefore, unsurprising that community adherence to self-guided CP is typically low. For example, the highest completion rate found in a recent review examining community usage of self-guided CP interventions for depression, anxiety, and mood enhancement was 19.5% [16], and the intervention that achieved this completion rate, CBT Psych, was actually an earlier version of Overcome Social Anxiety targeted toward stuttering populations. Other research has shown that for self-guided internet psychotherapy interventions, over 90% of users withdraw after only 2 sessions [39]. In comparison, less than one-third of users in our sample (112/369, 30.4%) withdrew before completing the prequestionnaires, module 1, and module 2, which together took

users an average of 2 hours and 13 min. It should be noted, however, that there is some experimental evidence suggesting that adding a financial cost to ICBT interventions increases adherence [32]. Some proportion of Overcome Social Anxiety's comparatively high adherence rate may, therefore, be attributable to its cost. Indeed, it would appear reasonable to expect that users who are willing to make financial sacrifices to access ICBT interventions are also more willing to sacrifice the time and effort necessary to complete those interventions, whereas those who are less committed may tend to opt for free ICBT instead.

Finally, user feedback was generally positive. For example, of the 35 completers who left feedback, 26 (26/35, 74%) left general praise and 18 (18/35, 51%) specifically stated that they had experienced a reduction in symptom severity. However, because feedback was solicited only at the end of the intervention, users who enjoyed and benefited from the intervention may be overrepresented, whereas users who did not enjoy or benefit from it may have tended to drop out before reaching the user feedback questions.

Limitations and Future Research

There remain a number of important questions for future research to address. First, although it has been shown that a therapist's assistance increases adherence to CCBT [38], therapist-assisted CP is not as scalable as self-guided CP. The further development of mechanisms to improve adherence to self-guided CP interventions remains an important avenue for future research. Second, it is a limitation of this study that we have no posttreatment data from noncompleters and therefore cannot report changes in their symptom severity following their partial completion of the intervention. Although past research shows that partial completion of ICBT interventions is beneficial [32], and some participants may indeed have ceased using the intervention after experiencing a satisfactory reduction in the severity of their symptoms, future research measuring symptoms periodically over the course of the treatment would be required to clarify the effects of partially completing Overcome Social Anxiety. Third, the intervention's apparent success cannot currently be attributed to any particular elements of its design. Overcome Social Anxiety was created to address 5 limitations of other ICBT interventions, but it remains unclear which of these limitations are most crucial for designers of future ICBT interventions to address. Fourth, although Overcome Social Anxiety has now been evaluated as an intervention for university undergraduates [14] and members of its general user base, future research would be required to evaluate the intervention in a clinical setting or to compare its effectiveness with human-delivered CBT. On a related note, although high FNE scores among our participants indicate that many of them may have met diagnostic criteria for social anxiety disorder, our lack of diagnostic interviews leaves us unable to draw any conclusions about the intervention's efficacy among those who did. Fifth, additional research exploring the predictors of program completion and symptom reduction could be useful in targeting the program toward those most likely to benefit from it or in establishing a screening mechanism to help prospective users decide whether it is an appropriate treatment for them. To this end, given our finding that completers had lower scores than noncompleters on the anxiety subscale of DASS, research examining that measure as a predictor of program outcomes could be considered. Sixth, the fact that the majority of users were men is interesting as a number of studies have shown that social anxiety is more prevalent among women [40]. Future research could explore the possibility that men value the privacy afforded by stand-alone ICBT programs more highly than women do. Finally, there is currently no data indicating whether and for what length of time users maintain reductions in symptom severity following their completion of the intervention.

Conclusions

Overcome Social Anxiety was initially evaluated through a randomized controlled trial, which indicated that the intervention reduces the severity of social anxiety symptoms among university undergraduates [14]. Although our study's lack of a control condition leaves us unable to draw causal inferences, we believe that it is reasonable to suppose that a considerable proportion of the pretreatment-to-posttreatment reduction in symptom severity may be attributable to the intervention, as this was a very large effect (Cohen d=1.76) and research shows that social anxiety tends to be persistent when it remains untreated [5]. Given this assumption, this study reinforces the findings of the initial trial [14] in 4 ways. First, it supports the finding that Overcome Social Anxiety is effective in reducing the severity of social anxiety symptoms. Second, its high external validity extends the initial trial's results to indicate that the intervention is highly effective among community users who complete it. Third, it suggests that the intervention's benefits are not limited to reducing the severity of social anxiety symptoms; the program appears to alleviate symptoms of depression, physiological symptoms of anxiety, and stress among its users as well. Finally, this study suggests that Overcome Social Anxiety has a high completion rate compared with other self-guided CP interventions. In summary, the results of this study converge with those of the initial trial's, attesting to Overcome Social Anxiety's effectiveness as a self-guided ICBT intervention and providing further indication that future interventions may be able to draw from elements of its design.

Acknowledgments

The authors would like to thank the users of Overcome Social Anxiety for making this study possible, Dr Neil Yager for his insightful comments and his help procuring program usage data, and Kirsten Gullickson for her assistance with coding user feedback data.

Conflicts of Interest

The authors FDH and RGM are cofounders of Overcome Social Anxiety. FDH is the director of AI-Therapy, the company that created Overcome Social Anxiety. Neither FDH nor RGM provided funding for the study or conducted any data analyses. All other authors declare no conflicts of interest.

Multimedia Appendix 1

Screenshot of Overcome Social Anxiety.

[PNG File, 897KB - jmir_v21i1e11566_app1.png]

References

1. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
2. Dell'Osso L, Rucci P, Ducci F, Ciapparelli A, Vivarelli L, Carlini M, et al. Social anxiety spectrum. Eur Arch Psychiatry Clin Neurosci 2003 Dec;253(6):286-291. [doi: 10.1007/s00406-003-0442-5] [Medline: 14714117]
3. Schneier FR, Heckelman LR, Garfinkel R, Campeas R, Fallon BA, Gitow A, et al. Functional impairment in social phobia. J Clin Psychiatry 1994 Aug;55(8):322-331. [Medline: 8071299]
4. Stein MB, Torgrud LJ, Walker JR. Social phobia symptoms, subtypes, and severity: findings from a community survey. Arch Gen Psychiatry 2000 Nov;57(11):1046-1052. [Medline: 11074870]
5. Reich J, Goldenberg I, Vasile R, Goisman R, Keller M. A prospective follow-along study of the course of social phobia. Psychiatry Res 1994 Dec;54(3):249-258. [Medline: 7792329]
6. Antony MM, Federici A, Stein MB. Overview and introduction to anxiety disorders. In: Oxford Handbook of Anxiety and Related Disorders. New York, NY, US: Oxford University Press; 2009:3-15.
7. Olfson M, Guardino M, Struening E, Schneier FR, Hellman F, Klein DF. Barriers to the treatment of social anxiety. Am J Psychiatry 2000 Apr;157(4):521-527. [Medline: 10739410]
8. Gould RA, Buckminster S, Pollack MH, Otto MW, Massachusetts LY. Cognitive-behavioral and pharmacological treatment for social phobia: a metaâ€analysis. Clin Psychol Sci Pract 1997;4:291-306. [doi: 10.1111/j.1468-2850.1997.tb00123.x]
9. Rodebaugh TL, Holaway RM, Heimberg RG. The treatment of social anxiety disorder. Clin Psychol Rev 2004 Nov;24(7):883-908. [doi: 10.1016/j.cpr.2004.07.007] [Medline: 15501560]
10. Kaltenthaler E, Cavanagh K. Computerised cognitive behavioural therapy and its uses. Progress Neurol Psychiatry 2010 Jun 09;14(3):22-29. [doi: 10.1002/pnp.163]
11. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
12. Helgadóttir F, Menzies RG, Onslow M, Packman A, O'Brian S. Online CBT II: A phase I trial of a standalone, online CBT treatment program for social anxiety in stuttering. Behav Change 2009 Dec 1;26(4):254-270. [doi: 10.1375/bech.26.4.254]
13. Adelman CB, Panza KE, Bartley CA, Bontempo A, Bloch MH. A meta-analysis of computerized cognitive-behavioral therapy for the treatment of DSM-5 anxiety disorders. J Clin Psychiatry 2014 Jul;75(7):e695-e704. [doi: 10.4088/JCP.13r08894] [Medline: 25093485]
14. McCall HC, Richardson CG, Helgadottir FD, Chen FS. Evaluating a web-based social anxiety intervention among university students: randomized controlled trial. J Med Internet Res 2018 Mar 21;20(3):e91 [FREE Full text] [doi: 10.2196/jmir.8630] [Medline: 29563078]
15. Helgadóttir F, Menzies RG, Onslow M, Packman A, O'Brian S. Online CBT I: bridging the gap between Eliza and modern online CBT treatment packages. Behav Change 2009 Dec 1;26(4):245-253. [doi: 10.1375/bech.26.4.245]
16. 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 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
17. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
18. Cavanagh K, Herbeck BB, Rothenberger SD, Abebe KZ, Rollman BL. My care manager, my computer therapy and me: the relationship triangle in computerized cognitive behavioural therapy. Internet Interv 2018 Mar;11:11-19 [FREE Full text] [doi: 10.1016/j.invent.2017.10.005] [Medline: 30135755]
19. Craske MG, Roy-Byrne P, Stein MB, Sullivan G, Hazlett-Stevens H, Bystritsky A, et al. CBT intensity and outcome for panic disorder in a primary care setting. Behav Ther 2006 Jun;37(2):112-119. [doi: 10.1016/j.beth.2005.05.003] [Medline: 16942966]
20. Wells A. The attention training technique: theory, effects, and a metacognitive hypothesis on auditory hallucinations. Cogn Behav Pract 2007 May;14(2):134-138. [doi: 10.1016/j.cbpra.2006.01.010]

21. Holmes EA, Arntz A, Smucker MR. Imagery rescripting in cognitive behaviour therapy: images, treatment techniques and outcomes. J Behav Ther Exp Psychiatry 2007 Dec;38(4):297-305. [doi: 10.1016/j.jbtep.2007.10.007] [Medline: 18035331]
22. Watson D, Friend R. Measurement of social-evaluative anxiety. J Consult Clin Psychol 1969 Aug;33(4):448-457. [Medline: 5810590]
23. Stopa L, Clark DM. Social phobia: comments on the viability and validity of an analogue research strategy and British norms for the fear of negative evaluation questionnaire. Behav Cogn Psychother 2001 Oct 23;29(4):423-430. [doi: 10.1017/S1352465801004039]
24. Leary MR. A brief version of the Fear of Negative Evaluation Scale. Pers Soc Psychol Bull 1983 Sep 1;9(3):371-375. [doi: 10.1177/0146167283093007]
25. Carleton RN, Collimore KC, McCabe RE, Antony MM. Addressing revisions to the Brief Fear of Negative Evaluation scale: measuring fear of negative evaluation across anxiety and mood disorders. J Anxiety Disord 2011 Aug;25(6):822-828. [doi: 10.1016/j.janxdis.2011.04.002] [Medline: 21565463]
26. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-U]
27. Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychol Assess 1998;10(2):176-181. [doi: 10.1037/1040-3590.10.2.176]
28. Crawford J, Henry JD. The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample. Br J Clin Psychol 2003 Jun;42(Pt 2):111-131. [doi: 10.1348/014466503321903544] [Medline: 12828802]
29. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
30. Creswell JW. Controversies in mixed methods research. In: Denzin NK, Lincoln YS, editors. The SAGE Handbook of Qualitative Research. Los Angeles, CA, US: SAGE Publications; 2011:269-284.
31. Unnebrink K, Windeler J. Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. Statist Med 2001 Dec 30;20(24):3931-3946. [doi: 10.1002/sim.1149]
32. Hilvert-Bruce Z, Rossouw PJ, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. Behav Res Ther 2012 Aug;50(7-8):463-468. [doi: 10.1016/j.brat.2012.04.001] [Medline: 22659155]
33. Stewart RE, Chambless DL. Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: a meta-analysis of effectiveness studies. J Consult Clin Psychol 2009 Aug;77(4):595-606. [doi: 10.1037/a0016032] [Medline: 19634954]
34. Vriends N, Bolt OC, Kunz SM. Social anxiety disorder, a lifelong disorder? A review of the spontaneous remission and its predictors. Acta Psychiatr Scand 2014 Aug;130(2):109-122. [doi: 10.1111/acps.12249] [Medline: 24506164]
35. Grant BF, Hasin DS, Blanco C, Stinson FS, Chou SP, Goldstein RB, et al. The epidemiology of social anxiety disorder in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. J Clin Psychiatry 2005 Nov;66(11):1351-1361. [Medline: 16420070]
36. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404]
37. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ, US: Lawrence Erlbaum Associates; 1988.
38. Titov N, Andrews G, Schwencke G, Mahoney A. Shyness 3: randomized controlled trial of guided versus unguided Internet-based CBT for social phobia. Aust N Z J Psychiatry 2008 Dec;42(12):1030-1040. [doi: 10.1080/00048670802512107]
39. Titov N, Dear BF, Johnston L, Lorian C, Zou J, Wootton B, et al. Improving adherence and clinical outcomes in self-guided internet treatment for anxiety and depression: randomised controlled trial. PLoS One 2013 Jul 3;8(7):e62873. [doi: 10.1371/journal.pone.0062873] [Medline: 23843932]
40. Asher M, Asnaani A, Aderka IM. Gender differences in social anxiety disorder: a review. Clin Psychol Rev 2017 Aug;56:1-12. [doi: 10.1016/j.cpr.2017.05.004] [Medline: 28578248]

Abbreviations

CBT: cognitive behavioral therapy
CCBT: computerized cognitive behavioral therapy
CP: computer-aided psychotherapy
DASS: Depression Anxiety Stress Scales
FNE: Fear of Negative Evaluation Scale
ICBT: internet-delivered cognitive behavioral therapy

Edited by G Eysenbach; submitted 13.07.18; peer-reviewed by F Talbot, K Stasiak; comments to author 30.08.18; revised version received 25.09.18; accepted 17.10.18; published 10.01.19.

Please cite as:
McCall HC, Helgadottir FD, Menzies RG, Hadjistavropoulos HD, Chen FS
Evaluating a Web-Based Social Anxiety Intervention Among Community Users: Analysis of Real-World Data
J Med Internet Res 2019;21(1):e11566
URL: http://www.jmir.org/2019/1/e11566/
doi:10.2196/11566
PMID:30632965

Original Paper

A Novel Insight Into the Challenges of Diagnosing Degenerative Cervical Myelopathy Using Web-Based Symptom Checkers

Benjamin Marshall Davies[1*], BSc (Hons), MBChB (Hons), MRCS; Colin Fraser Munro[2*], BA (Hons); Mark RN Kotter[1,3], MD, MPhil, PhD

[1]Academic Neurosurgery Unit, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom

[2]School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom

[3]Anne McLaren Laboratory for Regenerative Medicine, Wellcome–Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Mark RN Kotter, MD, MPhil, PhD
Academic Neurosurgery Unit
Department of Clinical Neurosciences
University of Cambridge
ED Adrian Building, Forvie Site
Robinson Way
Cambridge, CB2 2PY
United Kingdom
Phone: 44 7887790773
Email: mrk25@cam.ac.uk

Abstract

Background: Degenerative cervical myelopathy (DCM) is a common debilitating condition resulting from degeneration of the cervical spine. While decompressive surgery can halt disease progression, existing spinal cord damage is often permanent, leaving patients with lifelong disability. Early surgery improves the likelihood of recovery, yet the average time from the onset of symptoms to correct diagnosis is over 2 years. The majority of delays occur initially, before and within primary care, mainly due to a lack of recognition. Symptom checkers are widely used by patients before medical consultation and can be useful for preliminary triage and diagnosis. Lack of recognition of DCM by symptom checkers may contribute to the delay in diagnosis.

Objective: The aims of this study were to investigate whether Web-based symptom checkers were able to recognize relevant symptoms of DCM, to characterize the DCM differential they returned , and to evaluate the diagnostic performance of recognized DCM symptoms.

Methods: We pooled classical DCM symptoms from leading review articles. These symptoms were entered into the algorithms used by the top 20 symptom checker websites (N=4; Google Search). The most widely cited symptom checker, WebMD, was used to characterize the differential diagnosis for DCM symptoms.

Results: A total of 31 classical DCM symptoms were identified, of which 45% (14/31) listed DCM as a differential and 10% (3/31) placed DCM in the top third of the differential. The mean differential rank for motor symptoms was significantly better than that for arthritic symptoms (P=.01) and the average differential rank for all symptoms (P=.048). The symptom checker WebMD performed best at recognizing DCM, placing the condition nearer to the top of the differential list (mean rank of 5.6) than either Healthline (rank of 12.9, P=.02) or Healthtools.AARP (rank of 15.5, P=.001). On WebMD, only one combination of symptoms resulted in DCM as the primary differential: neck, shoulder, and arm pain with hand weakness. Moreover, 151 differential diagnoses for DCM symptoms were recorded on WebMD. Multiple sclerosis and peripheral neuropathy were the most common differentials, shortlisted for 52% (16/31) and 32% (10/31) of the DCM symptoms, respectively.

Conclusions: DCM symptoms are poorly identified by Web-based symptom checkers, which leads to a large differential of many other common conditions. While a diagnosis becomes more likely as the number of symptoms increases, this represents more advanced disease and will not support much-needed earlier diagnosis. Symptom checkers remain an attractive concept with potential. Further research is required to support their optimization.

(J Med Internet Res 2019;21(1):e10868) doi:10.2196/10868

KEYWORDS

cord compression; degenerative cervical myelopathy; diagnosis; differential; spondylosis

Introduction

Degenerative cervical myelopathy (DCM) is a debilitating and progressive condition that occurs when the cervical spinal cord is compressed by degenerative changes in surrounding structures. These degenerative changes, previously referred to as cervical spondylosis, include degeneration of intervertebral discs, osteophyte formation, ligamentous hypertrophy, spinal subluxation, and uncovertebral and facet joint hypertrophy [1,2].

The epidemiology of DCM is poorly characterized and has been reliant on "operative incidence" alone. This has contributed to misconceptions that it is a "rare" condition [3], whereas in fact it is estimated to be the most common spinal cord disorder [4]. For example, Kovalova et al identified that 59% (108/183) of patients from a randomly selected cohort of 40-80 year olds had magnetic resonance imaging signs of cervical cord compression and 1.1% (2/183) had undiagnosed DCM [5]. In the first prospective study of its kind, Bednarik et al showed that 8% of individuals with asymptomatic cord compression will develop DCM after 1 year and 22% over the total observation period (median follow-up, 44 months; range 2-12 years) [6]. Another more recent study has echoed these findings, with 10% of asymptomatic cord compression patients developing DCM at follow-up (median follow-up, 21 months; range 3-27 months) [7]. Given the association between DCM and age, as well as our aging population, its incidence is expected to rise.

Current treatment for DCM is limited to surgery that aims to relieve compression of the spinal cord. While most patients make a meaningful recovery, it is usually incomplete as existing damage is irreversible [8,9]. As such, treatment within 6 months of symptom onset has been shown to offer the best chance of making a full recovery [10,11].

Unfortunately, most patients wait much longer for a diagnosis; in the only study of its kind to date, Behrbalk et al found that the average time from onset of symptoms to correct diagnosis was 2.2 years [12]. Moreover, many patients go undiagnosed: in a series of neck of femur fracture patients, undiagnosed DCM was found in 18% of patients [13]. As a result, presently, most patients retain lifelong disabilities. This has a major socioeconomic impact on their lives, with unemployment, dependency, and a reduced quality of life. A recent study has demonstrated that patients with DCM have one of the lowest Short Form-36 scores among those with chronic diseases, lower than those with diabetes, cancer, chronic lung disease, and depression [14]. Thus, to improve the outcomes for such patients, an early diagnosis is needed.

Patients with DCM typically enter the health care system via primary care, and it is this interplay, between the onset of symptoms, patient presentation to primary care, consultation(s), and onward specialty referral, that makes up the majority of diagnostic delays [12]. The factors driving missed and delayed diagnosis are poorly characterized at present and difficult to investigate. However, the problem is likely multifactorial, including nonspecific and subtle early features often occurring in isolation initially, which may overlap with other conditions; incomplete neurological assessments by professionals; and poor awareness of the disease [1]. For example, in the Netherlands, a general practitioner is consulted 7 times a week for neck or upper extremity complaints (possible symptoms of DCM) of various causes [15], so distinguishing DCM can be difficult. In addition, the aforementioned Behrbalk et al series identified that 43% of patients with DCM were diagnosed and sometimes treated for carpal tunnel syndrome initially [12].

Web-based symptom checkers are websites that allow patients to select or enter a number of symptoms and using proprietary diagnostic algorithms produce a list of potential diagnoses, usually ranked in order of likelihood. These are popular with the general public; the leading engine, WebMD, receives 22 million unique visitors a month [16]. They are also frequently used prior to medical consultations; 45% of patients attending a genitourinary clinic, 47% of patients presenting to a rheumatology clinic, 53.5% of patients enrolled in a primary care practice, 52% of orthopedic outpatients, 51% of gastroenterology outpatients, 24% of adults accompanying children to a pediatric orthopedics clinic, 29% of patients referred to a medical genetics clinic, and 18% of otolaryngology outpatients with internet access had used the internet to research their symptoms prior to consultation [17-24]. Consequently, it is likely that symptom checkers are consulted when symptoms of DCM appear.

The results obtained from such websites may guide further searches and information seeking of patients. This may in turn influence their narration of symptoms and aid or obstruct diagnosis when they are seen in primary care.

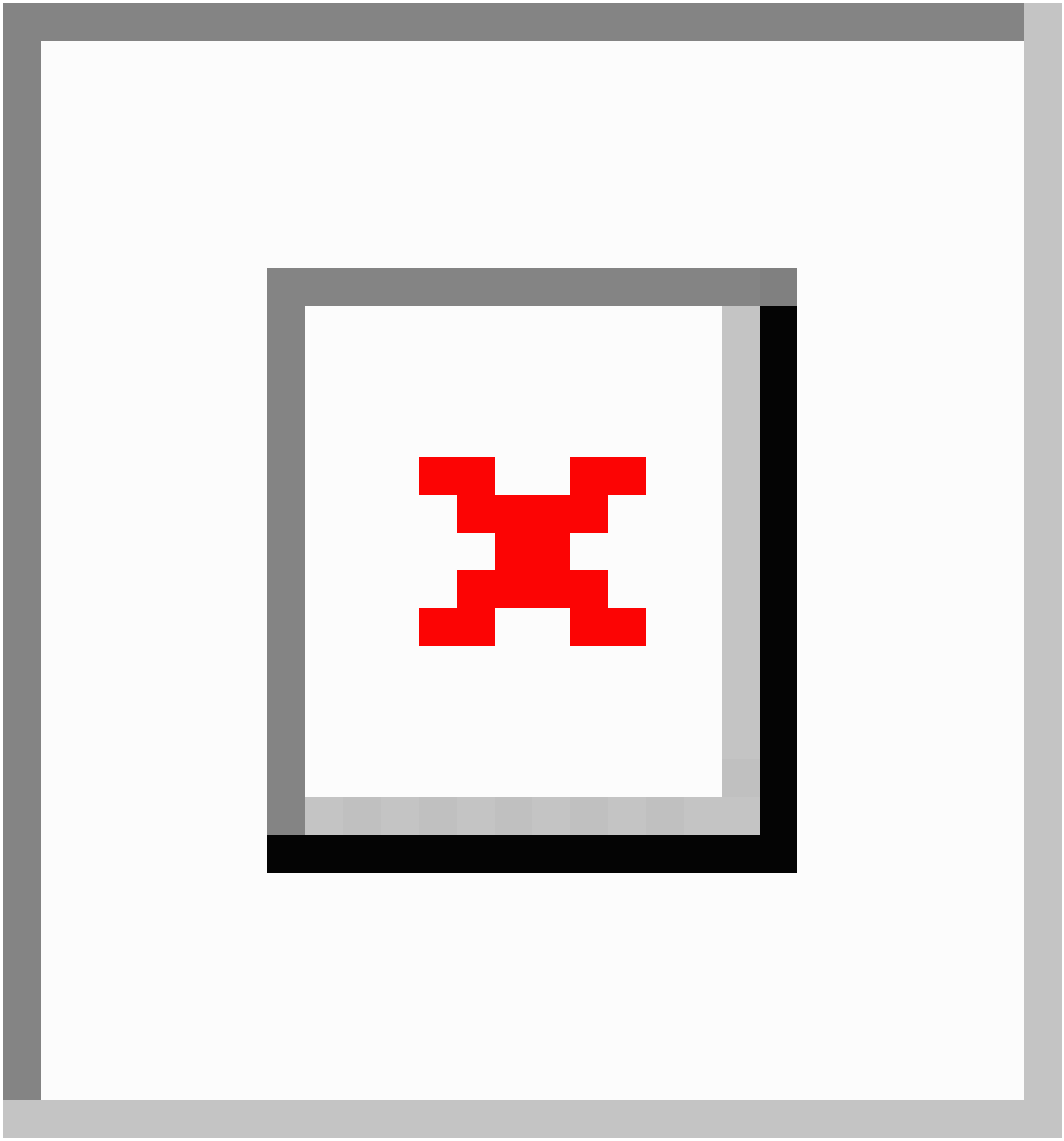
This study therefore sought to investigate the recognition of DCM in Web-based symptom checkers.

Methods

Reported Degenerative Cervical Myelopathy Symptoms

DCM symptoms were compiled from 4 leading review articles (cited by an average of 25 PubMed Central articles) published since 2000, taken from journals spanning a range of medical expertise: rehabilitation [25], neurology [2], neurosurgery [26], and primary care [27]. Reported symptoms were extracted from the articles and consolidated into a single list by removing duplicates or overlapping symptoms. In this article, these symptoms are referred to as "classical" symptoms.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram illustrating the selection of symptom checkers.



Web-Based Symptom Checkers

From a Google Search of "Symptom Checker" (returning over 4 million results), we selected the unique symptom checker search engines powering the top 20 symptom checkers. Checkers unable to make a diagnosis of DCM based upon symptoms or which did not list differential results in the order of likelihood were excluded (Figure 1). Consequently, we included the symptom checkers WebMD, Healthline, Healthtools.AARP, and NetDoctor. The estimated monthly visits for these websites were as follows: WebMD, 22 million [16]; Healthline, 11 million [28]; Healthtools.AARP, 3 million [29]; and NetDoctor, 1 million [30]. These symptom checker websites all described DCM under the umbrella term of (cervical) cord stenosis; however, for the purposes of this article, the term DCM will continue to be used.

Each literature-recognized DCM symptom was entered, either directly (NetDoctor) or by selecting the best match from a list of options (WebMD, Healthline, and Healthtools.AARP), into each symptom checker. When demographic information was required (WebMD and NetDoctor), the average age (57 years) and gender (male) from a recent DCM study were used [14]. If a region was required (NetDoctor), Western Europe was entered.

Analysis

The results of the symptom checkers were combined for analysis. Metrics of performance included (1) *differential rank* for DCM (the position of DCM in a list of differentials, eg,

fifth); (2) *total number of differentials* returned*;* and (3) *mean percentile rank* for DCM (percentage of conditions that were ranked below DCM in the differential list).

Additionally, WebMD, as the most widely cited and visited symptom checker (4 of the top 20 symptom checkers are powered by WebMD) with the best condition match for DCM, was selected to investigate the specificity of different combinations of symptoms and the overall differential for DCM.

To identify the most specific combination of symptoms, the symptoms that had yielded DCM as a differential were searched again in paired combinations using 2×2 probability tables, such that every combination was assessed. Symptom combinations that improved their mean ranking were carried forward into further 2×2 probability tables until no more improvements in differential rank could be achieved.

To identify the overall differential for DCM, we recorded the entire list of differential conditions for each search symptom. The number of times each of these conditions was a differential for any DCM symptom was recorded to allow a frequency chart to be plotted and the overall differential identified.

Statistics

The Shapiro-Wilk test was used to assess for parametric distribution of data sets. The Mann-Whitney *U* test was then used to compare the means of nonparametric distributions, while a two-tailed *t* test was used to compare the means of parametric distributions. The Pearson correlation coefficient was used to assess relationships between mean differential ranks, mean percentile ranks, and mean number of differentials.

Results

From the 4 review articles, 31 unique DCM symptoms were identified. These were grouped into motor, sensory, autonomic, and arthritic categories (Table 1) based on the domains of common outcome assessments [10]. Only abnormal gait, Lhermitte's sign, and urinary incontinence were listed by all 4 of the articles.

Accordingly, 45% (14/31) of the symptoms entered into the Web-based symptom checkers listed DCM as a possible diagnosis. No individual symptom placed DCM as a differential in all 4 Web-based symptom checkers or as the primary differential in any one (Table 1). Of the 3 ubiquitous DCM symptoms from the literature, abnormal gait and Lhermitte's sign did not yield DCM as a differential in the Web-based symptom checkers, and urinary incontinence had a mean rank of 15 out of 23 (percentile rank, 35). Of the symptoms which returned a differential of DCM, 14% (2/14) ranked DCM in the bottom third of differentials (percentile rank, 0-33.3), 64% (9/14) in the middle third of differentials (percentile rank, 33.3-66.6), and 21% (3/14) in the top third of differentials (percentile rank, 66.6-100). Upper limb or arm paresthesia, upper limb or arm pain (brachialgia), and hand paresthesia (with percentile ranks of 71.4, 66.7, and 70.6, respectively) were the only symptoms placing DCM in the top third of the differentials (Table 1).

The mean differential rank for all symptoms entered individually was 10.3, while that for the separate symptom categories were as follows: motor symptoms, 5.0; sensory symptoms, 9.8; autonomic symptoms, 14.5; and arthritic symptoms, 13.7 (Figure 2). A value of 1 would represent DCM as the top differential. The error bars indicate 95% CI. The N numbers for each category are indicated at the bottom of their respective bars.

The mean differential rank for motor symptoms was significantly better than that for arthritic symptoms (P=.01) and the average differential rank for all symptoms (P=.048). There were no other significant differences between the different symptom categories for the differential ranks, the number of differentials per symptom, or the percentile ranks.

Out of all the symptom checkers, WebMD placed DCM nearer to the top of the differential list (mean rank of 5.6) than either Healthline (rank of 12.9, P=.02) or Healthtools.AARP (rank of 15.5, P=.001). WebMD also returned fewer differential conditions for DCM symptoms (14.5) than Healthline (29.5, P=.01) and Healthtools.AARP (19.8, P=.0496). Unfortunately, comparisons with NetDoctor were not possible due to the low N number.

When symptoms were combined, on WebMD, the differential rank for DCM improved in the majority of circumstances. However, only a combination of neck pain, shoulder pain, upper limb or arm pain (brachialgia), and hand weakness placed DCM as the primary differential. There were 5 pairs of symptoms that gave DCM as the second differential and were as follows: (1) upper limb or arm weakness (paresis) and hand weakness; (2) upper limb or arm weakness (paresis) and hand numbness or sensory loss; (3) upper limb or arm weakness (paresis) and hand paresthesia; (4) hand weakness and hand numbness or sensory loss; (5) hand weakness and hand paresthesia.

For each of these pairs of symptoms, the condition ahead of DCM in the differential list was always peripheral neuropathy.

Overall, WebMD listed 151 differentials for DCM symptoms (Multimedia Appendix 1) with multiple sclerosis and peripheral neuropathy as the most common differentials (Figure 3). The list was collated by combining all the differential lists for each individual literature DCM symptom. DCM was listed for 10 symptoms and cervical spondylosis also for 10 symptoms. Carpal tunnel syndrome was only listed as a differential for DCM symptoms once (for clumsy hands).

The number of times a symptom was referenced in the literature did not differ between symptoms that identified and did not identify DCM. In fact, there was a trend between the number of times a symptom was referenced in the literature and the mean differential rank for DCM (R=0.49, P=.08), meaning that the symptoms referenced more often in the literature tended to rank DCM lower down the differential list.

Table 1. Degenerative cervical myelopathy (DCM) symptoms compiled from the 4 review articles (if DCM was given as a differential [italicized rows], the mean rank and percentile rank in the differential list was recorded).

DCM symptom[a]	Reviews in which symptom was mentioned, n (%)	Symptom checkers in which DCM was listed as a differential, n (%)	Mean differential rank for DCM	Mean total number of differentials	Mean percentile rank in differential list[b]
Motor symptoms					
Abnormal gait	4 (100)	0 (0)	N/A[c]	22	N/A
Loss of balance	*1 (25)*	*1 (25)*	*8*	*14*	*42.9*
General weakness	1 (25)	0 (0)	N/A	94	N/A
Lack of coordination	1 (25)	0 (0)	N/A	13	N/A
Upper limb or arm weakness (paresis)	*2 (50)*	*1 (25)*	*4*	*10*	*60*
Upper limb or arm spasticity	1 (25)	0 (0)	N/A	22	N/A
Clumsy hands	3 (75)	0 (0)	N/A	17	N/A
Loss of hand dexterity	1 (25)	0 (0)	N/A	6	N/A
Hand weakness	*1 (25)*	*1 (25)*	*4*	*5*	*20*
Lower limb or leg weakness (paresis)	3 (75)	0 (0)	N/A	14	N/A
Lower limb or leg spasticity	*1 (25)*	*2 (50)*	*5*	*19.5*	*61.2*
Lower limb or leg jerking	1 (25)	0 (0)	N/A	24	N/A
Lower limb or leg stiffness	2 (50)	0 (0)	N/A	15	N/A
Sensory symptoms					
Upper limb or arm numbness or sensory loss	*2 (50)*	*3 (75)*	*10*	*27.3*	*57.1*
Upper limb or arm paresthesia	*2 (50)*	*1 (25)*	*4*	*14*	*71.4*
Upper limb or arm pain (brachialgia)	*2 (50)*	*1 (25)*	*4*	*12*	*66.7*
Hand numbness or sensory loss	*2 (50)*	*3 (75)*	*12*	*22*	*40.7*
Hand paresthesia	*1 (25)*	*1 (25)*	*5*	*17*	*70.6*
Lower limb or leg numbness or sensory loss	*2 (50)*	*2 (50)*	*14*	*34*	*49.9*
Lower limb or leg paresthesia	1 (25)	0 (0)	N/A	30	N/A
Lhermitte's sign or phenomenon	4 (100)	0 (0)	N/A	3	N/A
Autonomic symptoms					
Fecal incontinence	3 (75)	0 (0)	N/A	16	N/A
Urgency of defecation	1 (25)	0 (0)	N/A	22	N/A
Urinary incontinence	*4 (100)*	*2 (50)*	*15*	*24.5*	*40.5*
Urinary urgency	2 (50)	0 (0)	N/A	18	N/A
Urinary frequency	1 (25)	0 (0)	N/A	28	N/A
Urinary hesitancy	1 (25)	0 (0)	N/A	13	N/A
Arthritic symptoms					
Neck stiffness	*2 (50)*	*1 (25)*	*7*	*12*	*41.7*
Neck pain	*3 (75)*	*3 (75)*	*15*	*21.7*	*33.0*
Neck crepitus or clicking	1 (25)	0 (0)	N/A	16	N/A
Shoulder pain	*1 (25)*	*3 (75)*	*15*	*27*	*43.4*

[a]For each symptom, the number of reviews mentioning it and the number of symptom checkers that give DCM as a differential for that symptom were

recorded.

[b]The mean percentile rank for DCM in the differential list represents the percentage of conditions that were ranked below DCM in the differential list (the higher the percentile rank, the more predictive the symptom). This allowed for comparison of DCM ranking among differential lists of differing lengths.

[c]N/A: not applicable.

Figure 2. The mean differential rank for all individual degenerative cervical myelopathy symptoms, as well as the mean differential rank of the individual symptoms grouped in the motor, sensory, autonomic, and arthritic categories. Asterisk denotes a statistically significant difference between motor and arthritic categories (*P*=.01); double asterisk denotes a statistically significant difference between motor and all categories (*P*=.048).

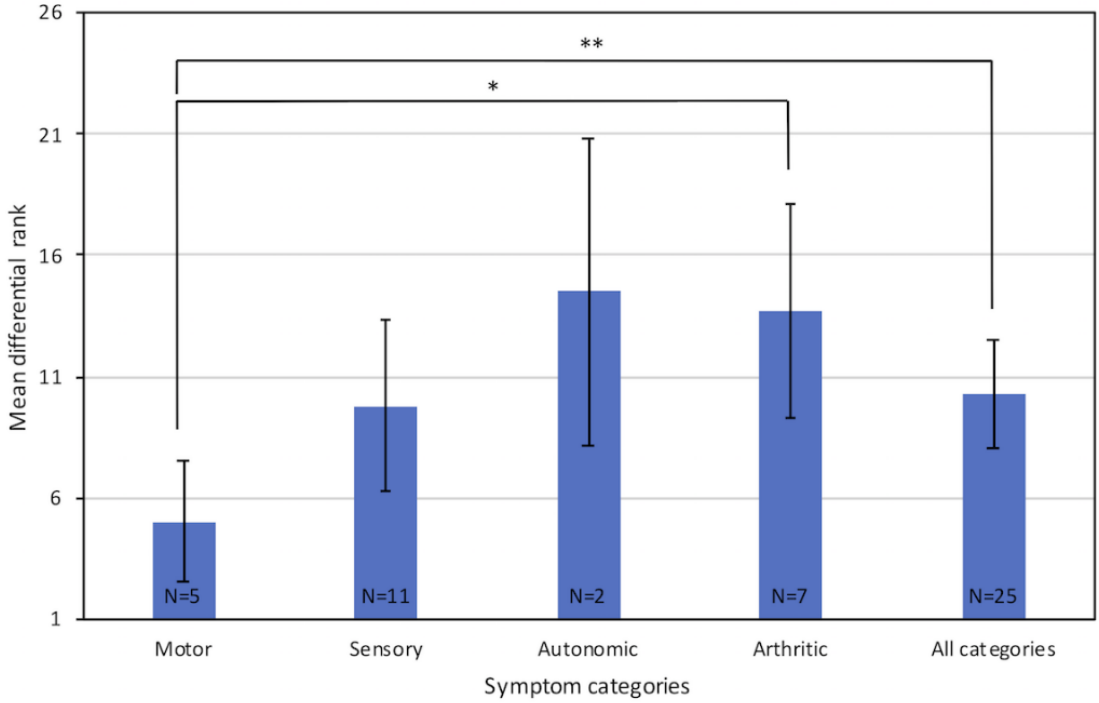
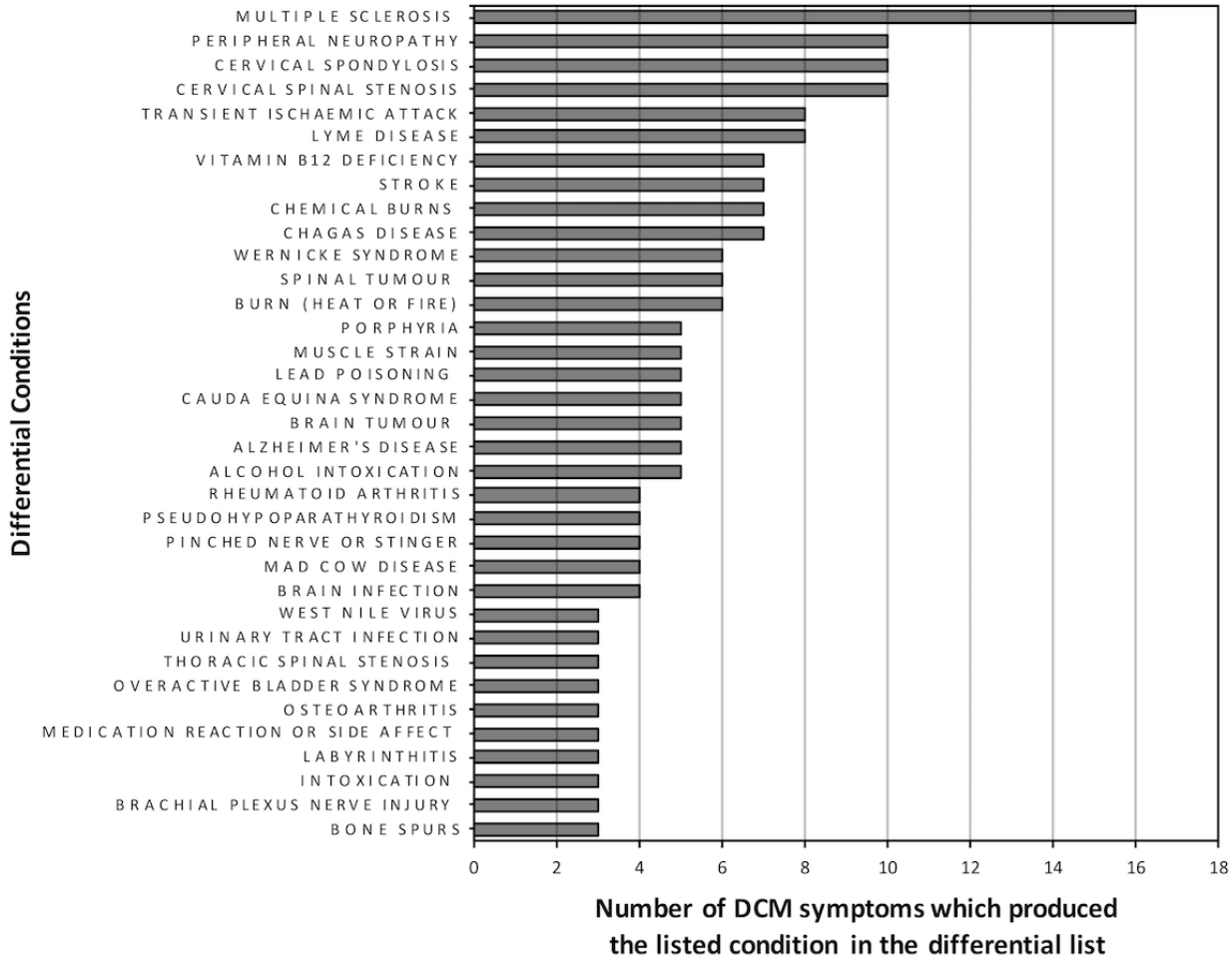


Figure 3. The number of degenerative cervical myelopathy (DCM) symptoms which produced the listed conditions in the differential list on WebMD.



Discussion

Principal Findings

Of each of the 31 DCM symptoms identified by the 4 review articles, only 45% (14/31) reported DCM as a differential in the Web-based symptom checkers. Additionally, key or prevalent literature symptoms fared no better. Of the symptoms identified by the symptom checkers, the majority (11/14, 79%) resulted in DCM being ranked in the bottom two-thirds of differentials. Multiple sclerosis and peripheral neuropathy were the most common differentials for DCM symptoms.

Therefore, based on the current classical descriptions of DCM, symptom checkers do not perform well at diagnosing the condition; moreover, if they did, DCM appeared toward the bottom of the differential list.

Can Symptom Checkers Have a Diagnostic Role in Degenerative Cervical Myelopathy?

Various studies have assessed the accuracy of Web-based symptom checkers with regard to linking symptoms with the correct diagnosis. For example, a recent study investigated the diagnostic accuracy of 23 symptom checkers, using 45 standardized patient vignettes, covering common and uncommon conditions (26 and 19 vignettes, respectively), as well as a range of triage urgencies (15 vignettes required emergency care, 15 required nonemergency care, and 15 required self-care) [31]. Each vignette was simplified into a core set of symptoms and entered into each symptom checker by an author with no clinical training. They found that the correct diagnosis was ranked first in 34% of patient cases and that the correct diagnosis was listed within the top 3 and top 20 differentials 51% and 58% of the time, respectively. Performance varied by the urgency of condition. The correct diagnosis was listed first for 24% of emergency cases, 38% of nonemergency cases, and 40% of self-care cases. Moreover, the correct diagnosis was listed first more often for common diagnoses than for uncommon diagnoses (38% vs 28%). Additional studies focusing on WebMD found that patients using the symptom checker in a hand surgery clinic correctly guessed a diagnosis matching that of the hand surgeon 33% of the time [32] and that with ear, nose, and throat patient cases, the symptom checker was correct 16% of the time, although the correct diagnosis was listed within the differential list 70% of the time [33].

Kobayashi et al had developed a screening questionnaire based solely on symptoms for the detection of DCM and demonstrated a sensitivity of 93.5% and specificity of 67.3% [34]. There are a number of limitations in this study, including its assessment on patients attending a neurosurgical clinic wherein the pretest probability will be greater than that in primary care, with more advanced and symptomatic patients with DCM, based on current practice. To our knowledge, this screening tool has not yet been

tested elsewhere and was unfortunately unable to be tested in this study due to the complex nature of several of the questions.

The diagnostic accuracy of symptom checkers in other fields and the successful development of a symptom screening tool by Kobayashi et al [34] suggest that their accuracy in DCM could improve. As proprietary tools, their algorithms could not be interrogated, but one assumes that their performance will be limited by the poorly populated diagnostic and epidemiological evidence base in DCM [1,4].

This is acceptable as supporting early diagnosis in DCM is a major research priority, particularly in primary care where the majority of diagnostic delays occur, and the growing popularity and penetrance of symptom checkers in public health-seeking behavior [17-24] suggests they are here to stay. However, the usability of symptom checkers may have some limitations. For instance, patients are required to have a reasonable level of language and computer proficiency in order to accurately input their symptoms, and the algorithms used may be less accurate in non-Western populations, where the prevalence of certain conditions may differ. Nevertheless, as single entities accessed by potential patients with some pretest probability and accessible by professionals as decision support tools, their optimization is more attractive as an intervention than standard alternatives, such as widespread education programs.

Apparent Degenerative Cervical Myelopathy Knowledge Gaps in Symptom Checkers

When considering the performance of symptom checkers with clinical practice, a number of deficiencies or knowledge gaps were identified, which if resolved, could optimize their diagnostic performance.

The symptom checkers generated a large differential for DCM, exemplifying the nonspecific nature of the symptoms and the diagnostic challenge. Of these differentials, alongside multiple sclerosis, peripheral neuropathy was predominant. In the only study so far to consider diagnostic practice in DCM, Behrbalk et al identified that 43% of patients were initially misdiagnosed with carpal tunnel syndrome [12]. The distinction of DCM from neuropathy is clearly a common pitfall.

The focus of symptom checkers is skewed to the upper limb, with neurological dysfunction. Furthermore, pain is a key symptom for the detection of DCM, particularly upper limb pain combined with neck and shoulder pain, which was required to promote DCM to the top of the differential list in WebMD. This is an interesting finding, as pain is neither always present in DCM nor an indication for treatment [35] and in fact often not recorded in many DCM clinical trials [10,36]. It is possible that this is a result of cervical radiculopathy, of which pain is typically a feature, being included by the term “cord stenosis” on the symptoms checkers. However, conversely focusing on diagnostic practice, Mizer et al found that upper limb pain had the best diagnostic odds ratio for DCM, with a value of 29.00, out of a long list of symptoms [37]. Therefore, while pain has not been a focus for treatment research, its significance in early diagnosis requires further evaluation.

DCM also affects the lower limb and autonomic nervous system, yet these symptoms demonstrated poor diagnostic utility in symptom checkers. This is significant, as some recent work suggests that lower limb symptoms and signs may in fact be the earliest clinical presentation of DCM, with gait disturbance being the most frequently presenting symptom in a prospective study of patients with initially nonmyelopathic cervical cord compression [38].

DCM only comes to the forefront of a differential diagnosis when multiple symptoms are reported. Unfortunately, this is likely to reflect more advanced disease, as assessed by the modified Japanese Orthopaedic Association scale and signs of cord compression on magnetic resonance imaging [39-42]; further research is required for early disease detection.

Potential Directions for Optimization of Symptom Checkers in Degenerative Cervical Myelopathy

Clearly generating the necessary epidemiological and early presentation data confounding standard clinical practice is likely to be helpful. However, this is not straightforward, and its current absence more likely reflects the difficult practicalities of conducting such research [4]. However, the lack of specificity of classical DCM symptoms may always be a limitation, and avenues to improve this would also be helpful.

An interesting finding from the aforementioned Kobayashi et al study was the predictive power of symptoms we would consider “atypical,” for example, chest tightness. This was not identified in this study as a “classical” symptom, yet in their series, the odds ratio of chest tightness in myelopathy patients compared with controls was 22.9 [34]. Other atypical symptoms reported as prevalent among patients with DCM include the following: respiratory dysfunction, with a reduction in both forced vital capacity and forced expiratory volume in 1 second in patients with DCM compared with that in controls (65% vs 88% and 72% vs 96%, respectively) [43]; hypertension (47% vs 30%, respectively) [44]; and headaches (88% of patients with DCM undergoing surgery) [45]. In these studies, all of these symptoms improved with surgical treatment of DCM, and none of these symptoms are typical of many of the identified DCM differentials. The prevalence of atypical symptoms therefore warrants more attention, as their combination with “classical” symptoms might help improve the specificity of screening tools.

Clinical examination is an important component of medical diagnosis. Moreover, it may be that symptoms alone are insufficient for screening of DCM and physical examination findings could help. This is not straightforward as, similar to symptoms, examination findings are inconsistent. Rhee et al found that although myelopathic physical signs were significantly more common in patients with DCM than in controls with neck-related complaints (79% vs 57%, respectively), the signs were not highly sensitive for diagnosing DCM, with 21% of patients with DCM having no myelopathic signs at all [46]. However, similar to symptoms, combinations of examination findings may improve diagnostic accuracy. Tejus et al showed that a combination of the finger flexion, Hoffman’s reflex, and plantar reflex could be effectively used as a marker of cervical spinal cord compression in patients with neck-related complaints, with a sensitivity of 91.7% and a specificity of 87.5%, and their absence had a negative predictive value of 77.8% [47]. While examination findings would be limited to

professional use, they could be a helpful addition to diagnostic algorithms, and their predictive power with symptoms requires further assessment.

Conclusions

Classical DCM symptoms perform poorly in Web-based symptom checkers, in particular lower limb and autonomic symptoms. While combinations of symptoms improve the diagnostic accuracy, this will not be useful for early diagnosis. With over 150 potential differentials listed, detecting DCM early is difficult. The development of accurate diagnostic strategies is needed to support earlier diagnosis and improve patient outcomes. Symptom checkers remain an attractive concept with potential. Further research in this area is required.

Acknowledgments

Research in the laboratory of MRNK is supported by a core support grant from the Wellcome Trust and Medical Research Council to the Wellcome–Medical Research Council Cambridge Stem Cell Institute. MRNK is supported by a National Institute for Health Research Clinician Scientist Award.

This report is independent research arising from a Clinician Scientist Award, CS-2015-15-023, supported by the National Institute for Health Research. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, or the Department of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

An alphabetical list of all the potential differential conditions given for literature degenerative cervical myelopathy symptoms by WebMD.

[PDF File (Adobe PDF File), 23KB - jmir_v21i1e10868_app1.pdf]

References

1. Davies BM, Mowforth OD, Smith EK, Kotter MR. Degenerative cervical myelopathy. BMJ 2018 Dec 22;360:k186 [FREE Full text] [Medline: 29472200]
2. Tracy JA, Bartleson JD. Cervical spondylotic myelopathy. Neurologist 2010 May;16(3):176-187. [doi: 10.1097/NRL.0b013e3181da3a29] [Medline: 20445427]
3. McCartney S, Baskerville R, Blagg S, McCartney D. Cervical radiculopathy and cervical myelopathy: diagnosis and management in primary care. Br J Gen Pract 2018 Jan;68(666):44-46. [doi: 10.3399/bjgp17X694361] [Medline: 29284641]
4. Nouri A, Tetreault L, Singh A, Karadimas SK, Fehlings MG. Degenerative Cervical Myelopathy: Epidemiology, Genetics, and Pathogenesis. Spine (Phila Pa 1976) 2015 Jun 15;40(12):E675-E693. [doi: 10.1097/BRS.0000000000000913] [Medline: 25839387]
5. Kovalova I, Kerkovsky M, Kadanka Z, Kadanka Z, Nemec M, Jurova B, et al. Prevalence and Imaging Characteristics of Nonmyelopathic and Myelopathic Spondylotic Cervical Cord Compression. Spine (Phila Pa 1976) 2016 Dec 15;41(24):1908-1916. [doi: 10.1097/BRS.0000000000001842] [Medline: 27509189]
6. Bednarik J, Kadanka Z, Dusek L, Kerkovsky M, Vohanka S, Novotny O, et al. Presymptomatic spondylotic cervical myelopathy: an updated predictive model. Eur Spine J 2008 Mar;17(3):421-431 [FREE Full text] [doi: 10.1007/s00586-008-0585-1] [Medline: 18193301]
7. Martin AR, De Leener B, Cohen-Adad J, Cadotte DW, Nouri A, Wilson JR, et al. Can microstructural MRI detect subclinical tissue injury in subjects with asymptomatic cervical spinal cord compression? A prospective cohort study. BMJ Open 2018 Dec 13;8(4):e019809 [FREE Full text] [doi: 10.1136/bmjopen-2017-019809] [Medline: 29654015]
8. Fehlings MG, Wilson JR, Kopjar B, Yoon ST, Arnold PM, Massicotte EM, et al. Efficacy and safety of surgical decompression in patients with cervical spondylotic myelopathy: results of the AOSpine North America prospective multi-center study. J Bone Joint Surg Am 2013 Sep 18;95(18):1651-1658. [doi: 10.2106/JBJS.L.00589] [Medline: 24048552]
9. Fehlings MG, Ibrahim A, Tetreault L, Albanese V, Alvarado M, Arnold P, et al. A global perspective on the outcomes of surgical decompression in patients with cervical spondylotic myelopathy: results from the prospective multicenter AOSpine international study on 479 patients. Spine (Phila Pa 1976) 2015 Sep 01;40(17):1322-1328. [doi: 10.1097/BRS.0000000000000988] [Medline: 26020847]
10. Davies BM, McHugh M, Elgheriani A, Kolias AG, Tetreault L, Hutchinson PJA, et al. The reporting of study and population characteristics in degenerative cervical myelopathy: A systematic review. PLoS One 2017;12(3):e0172564 [FREE Full text] [doi: 10.1371/journal.pone.0172564] [Medline: 28249017]

11. Tetreault LA, Karpova A, Fehlings MG. Predictors of outcome in patients with degenerative cervical spondylotic myelopathy undergoing surgical treatment: results of a systematic review. Eur Spine J 2015 Apr;24 Suppl 2:236-251. [doi: 10.1007/s00586-013-2658-z] [Medline: 23386279]
12. Behrbalk E, Salame K, Regev GJ, Keynan O, Boszczyk B, Lidar Z. Delayed diagnosis of cervical spondylotic myelopathy by primary care physicians. Neurosurg Focus 2013 Jul;35(1):E1. [doi: 10.3171/2013.3.FOCUS1374] [Medline: 23815245]
13. Radcliff KE, Curry EP, Trimba R, Walker JB, Purtill JJ, Austin MS, et al. High Incidence of Undiagnosed Cervical Myelopathy in Patients With Hip Fracture Compared With Controls. J Orthop Trauma 2016 Apr;30(4):189-193. [doi: 10.1097/BOT.0000000000000485] [Medline: 26562581]
14. Oh T, Lafage R, Lafage V, Protopsaltis T, Challier V, Shaffrey C, et al. Comparing Quality of Life in Cervical Spondylotic Myelopathy with Other Chronic Debilitating Diseases Using the Short Form Survey 36-Health Survey. World Neurosurg 2017 Oct;106:699-706. [doi: 10.1016/j.wneu.2016.12.124] [Medline: 28065875]
15. Bot SDM, van der Waal JM, Terwee CB, van der Windt DAWM, Schellevis FG, Bouter LM, et al. Incidence and prevalence of complaints of the neck and upper extremity in general practice. Ann Rheum Dis 2005 Jan;64(1):118-123 [FREE Full text] [doi: 10.1136/ard.2003.019349] [Medline: 15608309]
16. TrafficEstimate.com. 2017. symptoms.webmd.com Website Traffic and Information URL: https://www.trafficestimate.com/symptoms.webmd.com [accessed 2017-11-07] [WebCite Cache ID 6umJ0fk4C]
17. Christian SM, Kieffer SA, Leonard NJ. Medical genetics and patient use of the Internet. Clin Genet 2001 Sep;60(3):232-236. [Medline: 11595026]
18. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the Internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [Medline: 11929503]
19. Gupte CM, Hassan ANA, McDermott ID, Thomas RD. The internet--friend or foe? A questionnaire study of orthopaedic out-patients. Ann R Coll Surg Engl 2002 May;84(3):187-192 [FREE Full text] [Medline: 12092873]
20. O'Connor JB, Johanson JF. Use of the Web for medical information by a gastroenterology clinic population. JAMA 2000 Oct 18;284(15):1962-1964. [Medline: 11035893]
21. Beall MS, Golladay GJ, Greenfield MLVH, Hensinger RN, Biermann JS. Use of the Internet by pediatric orthopaedic outpatients. J Pediatr Orthop 2002;22(2):261-264. [Medline: 11856943]
22. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? BMC Musculoskelet Disord 2016 Dec 24;17(1):362 [FREE Full text] [doi: 10.1186/s12891-016-1189-2] [Medline: 27553253]
23. Schembri G, Schober P. The Internet as a diagnostic aid: the patients' perspective. Int J STD AIDS 2009 Apr;20(4):231-233. [doi: 10.1258/ijsa.2008.008339] [Medline: 19304965]
24. Tassone P, Georgalas C, Patel NN, Appleby E, Kotecha B. Do otolaryngology out-patients use the internet prior to attending their appointment? J Laryngol Otol 2004 Jan;118(1):34-38. [doi: 10.1258/002221504322731600] [Medline: 14979970]
25. Kalsi-Ryan S, Karadimas SK, Fehlings MG. Cervical spondylotic myelopathy: the clinical phenomenon and the current pathobiology of an increasingly prevalent and devastating disorder. Neuroscientist 2013 Aug;19(4):409-421. [doi: 10.1177/1073858412467377] [Medline: 23204243]
26. Amenta PS, Ghobrial GM, Krespan K, Nguyen P, Ali M, Harrop JS. Cervical spondylotic myelopathy in the young adult: a review of the literature and clinical diagnostic criteria in an uncommon demographic. Clin Neurol Neurosurg 2014 May;120:68-72. [doi: 10.1016/j.clineuro.2014.02.019] [Medline: 24731579]
27. Young WF. Cervical spondylotic myelopathy: a common cause of spinal cord dysfunction in older persons. Am Fam Physician 2000 Sep 01;62(5):1064-70, 1073 [FREE Full text] [Medline: 10997531]
28. TrafficEstimate.com. 2017. Healthline Website, Traffic and Information URL: https://www.trafficestimate.com/healthline.com [accessed 2017-11-07] [WebCite Cache ID 6umJMqvZL]
29. TrafficEstimate.com. 2017. Healthtools Website, Traffic and Information URL: https://www.trafficestimate.com/healthtools.aarp.org [accessed 2017-11-07] [WebCite Cache ID 6umJdzgZJ]
30. TrafficEstimate.com. 2017. NetDoctor Website, Traffic and Information URL: https://www.trafficestimate.com/netdoctor.co.uk [accessed 2017-11-07] [WebCite Cache ID 6umJoKGF7]
31. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [Medline: 26157077]
32. Hageman MGJS, Anderson J, Blok R, Bossen JKJ, Ring D. Internet self-diagnosis in hand surgery. Hand (N Y) 2015 Sep;10(3):565-569 [FREE Full text] [doi: 10.1007/s11552-014-9707-x] [Medline: 26330798]
33. Farmer SEJ, Bernardotto M, Singh V. How good is Internet self-diagnosis of ENT symptoms using Boots WebMD symptom checker? Clin Otolaryngol 2011 Oct;36(5):517-518. [doi: 10.1111/j.1749-4486.2011.02375.x] [Medline: 22032458]
34. Kobayashi H, Kikuchi S, Otani K, Sekiguchi M, Sekiguchi Y, Konno S. Development of a self-administered questionnaire to screen patients for cervical myelopathy. BMC Musculoskelet Disord 2010 Nov 22;11:268 [FREE Full text] [doi: 10.1186/1471-2474-11-268] [Medline: 21092213]
35. Fehlings M, Tetreault L, Riew K, Middleton J, Aarabi B, Arnold P, et al. A Clinical Practice Guideline for the Management of Patients With Degenerative Cervical Myelopathy: Recommendations for Patients With Mild, Moderate, and Severe Disease and Nonmyelopathic Patients With Evidence of Cord Compression. Global Spine J 2017 Sep;7(3 Suppl):70S-83S [FREE Full text] [doi: 10.1177/2192568217701914] [Medline: 29164035]

36. Davies BM, McHugh M, Elgheriani A, Kolias AG, Tetreault LA, Hutchinson PJA, et al. Reported Outcome Measures in Degenerative Cervical Myelopathy: A Systematic Review. PLoS One 2016;11(8):e0157263 [FREE Full text] [doi: 10.1371/journal.pone.0157263] [Medline: 27482710]
37. Mizer A, Bachmann A, Gibson J, Donaldson MB. Self-report and subjective history in the diagnosis of painful neck conditions: A systematic review of diagnostic accuracy studies. Musculoskelet Sci Pract 2017 Oct;31:30-44. [doi: 10.1016/j.msksp.2017.06.002] [Medline: 28644963]
38. Kadanka Z, Adamova B, Kerkovsky M, Kadanka Z, Dusek L, Jurova B, et al. Predictors of symptomatic myelopathy in degenerative cervical spinal cord compression. Brain Behav 2017 Sep;7(9):e00797 [FREE Full text] [doi: 10.1002/brb3.797] [Medline: 28948090]
39. Karpova A, Arun R, Kalsi-Ryan S, Massicotte EM, Kopjar B, Fehlings MG. Do quantitative magnetic resonance imaging parameters correlate with the clinical presentation and functional outcomes after surgery in cervical spondylotic myelopathy? A prospective multicenter study. Spine (Phila Pa 1976) 2014 Aug 15;39(18):1488-1497. [doi: 10.1097/BRS.0000000000000436] [Medline: 24859570]
40. Sadasivan KK, Reddy RP, Albright JA. The natural history of cervical spondylotic myelopathy. Yale J Biol Med 1993;66(3):235-242 [FREE Full text] [Medline: 8209559]
41. Tetreault L, Kopjar B, Nouri A, Arnold P, Barbagallo G, Bartels R, et al. The modified Japanese Orthopaedic Association scale: establishing criteria for mild, moderate and severe impairment in patients with degenerative cervical myelopathy. Eur Spine J 2017 Dec;26(1):78-84. [doi: 10.1007/s00586-016-4660-8] [Medline: 27342612]
42. Nouri A, Martin AR, Kato S, Reihani-Kermani H, Riehm LE, Fehlings MG. The Relationship Between MRI Signal Intensity Changes, Clinical Presentation, and Surgical Outcome in Degenerative Cervical Myelopathy: Analysis of a Global Cohort. Spine (Phila Pa 1976) 2017 Dec 15;42(24):1851-1858. [doi: 10.1097/BRS.0000000000002234] [Medline: 28498290]
43. Bhagavatula ID, Bhat DI, Sasidharan GM, Mishra RK, Maste PS, Vilanilam GC, et al. Subclinical respiratory dysfunction in chronic cervical cord compression: a pulmonary function test correlation. Neurosurg Focus 2016 Jun;40(6):E3. [doi: 10.3171/2016.3.FOCUS1647] [Medline: 27246486]
44. Li Z, Zhao Y, Jia W, Wang X, Chen B, Shahbaz M, et al. Surgical Treatment of Cervical Spondylotic Myelopathy Associated Hypertension--A Retrospective Study of 309 Patients. PLoS One 2015;10(7):e0133828 [FREE Full text] [doi: 10.1371/journal.pone.0133828] [Medline: 26193469]
45. Schrot RJ, Mathew JS, Li Y, Beckett L, Bae HW, Kim KD. Headache relief after anterior cervical discectomy: post hoc analysis of a randomized investigational device exemption trial: clinical article. J Neurosurg Spine 2014 Aug;21(2):217-222 [FREE Full text] [doi: 10.3171/2014.4.SPINE13669] [Medline: 24836655]
46. Rhee JM, Heflin JA, Hamasaki T, Freedman B. Prevalence of physical signs in cervical myelopathy: a prospective, controlled study. Spine (Phila Pa 1976) 2009 Apr 20;34(9):890-895. [doi: 10.1097/BRS.0b013e31819c944b] [Medline: 19352222]
47. Tejus MN, Singh V, Ramesh A, Kumar VRR, Maurya VP, Madhugiri VS. An evaluation of the finger flexion, Hoffman's and plantar reflexes as markers of cervical spinal cord compression - A comparative clinical study. Clin Neurol Neurosurg 2015 Jul;134:12-16. [doi: 10.1016/j.clineuro.2015.04.009] [Medline: 25909467]

Abbreviations

DCM: degenerative cervical myelopathy

Edited by G Eysenbach; submitted 25.04.18; peer-reviewed by A Nouri, H Kobayashi; comments to author 31.05.18; revised version received 23.07.18; accepted 30.07.18; published 11.01.19.

Please cite as:
Davies BM, Munro CF, Kotter MRN
A Novel Insight Into the Challenges of Diagnosing Degenerative Cervical Myelopathy Using Web-Based Symptom Checkers
J Med Internet Res 2019;21(1):e10868
URL: http://www.jmir.org/2019/1/e10868/
doi:10.2196/10868
PMID:30300137

Review

Internet-Delivered Acceptance and Commitment Therapy for Anxiety Treatment: Systematic Review

Joshua Kelson[1], PhD; Audrey Rollin[1], BS, MS; Brad Ridout[1], PhD; Andrew Campbell[1], PhD

Cyberpsychology Research Group, Faculty of Health Sciences, The University of Sydney, Sydney, Australia

Corresponding Author:
Joshua Kelson, PhD
Cyberpsychology Research Group
Faculty of Health Sciences
The University of Sydney
City Road
Sydney, 2006
Australia
Phone: 61 2 9351 7089
Email: joshnkelson@gmail.com

Abstract

Background: Anxiety conditions are debilitating and prevalent throughout the world. Acceptance and Commitment Therapy (ACT) is an effective, acceptance-based behavioral therapy for anxiety. However, there are treatment barriers (eg, financial, geographical, and attitudinal), which prevent people from accessing it. To overcome these barriers, internet-delivered ACT (iACT) interventions have been developed in recent years. These interventions use websites to deliver ACT information and skill training exercises on the Web, either as pure self-help or with therapist guidance.

Objective: This systematic review aimed to examine the therapeutic impact of iACT on all anxiety conditions.

Methods: The EMBASE, MEDLINE, ProQuest Central, PsycINFO, Scopus, and Web of Science databases were searched up to September 2018. The titles and abstracts of remaining records after deduplication were screened by 2 authors with a total of 36 full-text articles being retained for closer inspection next to eligibility criteria. Empirical studies of all designs, population types, and comparator groups were included if they appraised the impact of iACT treatment on any standardized measure of anxiety. Included studies were appraised on methodological quality and had their data extracted into a standardized coding sheet. Findings were then tabulated, and a narrative synthesis was performed because of the heterogeneity found between studies.

Results: A total of 20 studies met inclusion criteria. There were 11 randomized controlled trials (RCTs) and 9 uncontrolled pilot studies. Participants across all studies were adults. The anxiety conditions treated were as follows: generalized anxiety disorder (GAD), social anxiety disorder (SAD), illness anxiety disorder (IAD), and general anxiety symptoms, with or without comorbid physical and mental health problems. A total of 18 studies reported significant anxiety reduction after iACT treatment. This was observed in studies that delivered iACT with (n=13) or without (n=5) therapist guidance. The average attrition rate across all included studies during the active iACT treatment phase was 19.19%. In the 13 studies that assessed treatment satisfaction, participants on average rated their iACT experience with above average to high treatment satisfaction.

Conclusions: These findings indicate that iACT can be an efficacious and acceptable treatment for adults with GAD and general anxiety symptoms. More RCT studies are needed to corroborate these early iACT findings using empirical treatments in active control groups (eg, internet-delivered cognitive behavioral therapy). This would potentially validate the promising results found for SAD and IAD as well as address the full spectrum of anxiety disorders.

(J Med Internet Res 2019;21(1):e12530) doi:10.2196/12530

KEYWORDS

anxiety; anxiety disorders; acceptance and commitment therapy; mindfulness; telemedicine; internet; e-therapy

Introduction

Background

Anxiety and its related conditions are highly prevalent on a global scale [1]. The *Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition* (DSM-5) describes multiple anxiety disorders such as specific phobia, generalized anxiety disorder (GAD), social anxiety disorder (SAD), illness anxiety disorder (IAD), panic disorder, agoraphobia, separation anxiety disorder, substance or medication-induced anxiety disorder, and selective mutism [2]. These disorders can start early in life [3], persist long after onset [4], and be highly debilitating from even subclinical symptomatology [5,6]. They are often comorbid with other mental health issues [7], and risk factors for suicide [8].

A potential solution for people with anxiety is Acceptance and Commitment Therapy (ACT)—a psychological treatment that teaches mindfulness skills to help people accept their anxiety and commit to living in accordance with personal values [9]. The primary aim of ACT is not to eliminate anxiety symptoms but rather to improve psychological flexibility which refers to a person's ability to contact the present moment more fully as a conscious human being, and engage in values-based action [9,10]. This is accomplished in therapy by targeting 6 core processes of change: acceptance, contact with the present moment, cognitive defusion, self-as-context, values, and committed action [9]. ACT delivery is not constrained to one specific method or technology, but can make use of a host of methods and modalities to facilitate these processes [10]. Even though anxiety symptom reduction is not the primary focus of ACT, substantial evidence has accrued showing that it can significantly reduce anxiety symptoms and put anxiety disorders into remission [11-15]. In contrast to other established treatments for anxiety, such as cognitive behavioral therapy (CBT) [16,17] and pharmacotherapy [18], the ACT model provides a unique transdiagnostic approach [19] that emphasizes acceptance, mindfulness, and values-guided behavioral exercises, rather than control, logical analysis, and cognitive disputation exercises [10].

To make psychological treatment accessible to the broader population, researchers have developed Web-based therapy interventions [20]. This therapy involves using websites to provide mental health information and skill training online [21]. These interventions can overcome many of the structural and attitudinal barriers of face-to-face treatment such as therapist worker shortages, geographical distance, wait-lists, social stigma, financial cost, and work commitments [22], which are especially pronounced in rural areas [23]. Evidence indicates that Web-based interventions can be an effective, an acceptable, and a practical health care option for anxiety sufferers when used as pure self-help or as an adjunct to treatment-as-usual [24,25]. However, efforts to treat anxiety through the Web have largely been concentrated on CBT [26-29]. This is because the CBT tradition has a larger evidence base from a longer research history than the more recent ACT model [15,30]. Nevertheless, attrition and treatment adherence remain key issues for internet-delivered CBT (iCBT) [31,32]. It is therefore important to assess the potential benefits of other therapy models delivered through the Web, such as ACT. A previous review of randomized controlled trial (RCT) evidence indicated that internet-delivered ACT (iACT) can reduce anxiety and depression symptoms [33]. However, none of the RCTs published in this earlier review specifically targeted anxiety disorders nor examined the use of iACT for anxiety in other research designs.

Objective

The aim of this systematic review was to provide a comprehensive and up-to-date account on the empirical status of iACT for anxiety. To this end, all studies that investigated the impact of iACT on any standardized measure of anxiety were appraised. This included pilot studies as they can cost-effectively determine the feasibility of new interventions and research pursuits before expending greater resources on developing complex interventions, or testing advanced research designs [34]. As iACT for anxiety is a nascent research field [33], including the following information can help: (1) inform researchers on what has already been done, and where treatment gaps for anxiety are present; (2) clarify what interventions hold the potential for acceptable, efficacious, and practical anxiety treatment; and (3) potentially guide the design, development, and empirical validation of new iACT interventions as either self-help or as an adjunct to treatment-as-usual.

Methods

Eligibility Criteria

This systematic review was performed using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. A PRISMA checklist is available in Multimedia Appendix 1. To be included in the review, articles needed to report on empirical data obtained from iACT treatment on any standardized measure of anxiety. This criterion was set to ensure the validity and reliability of the anxiety symptom scores obtained at pre- and post-iACT treatment. Furthermore, to ensure that the review was comprehensive, all empirical research designs were included, as well as all population and comparator types. The iACT intervention under investigation needed to be delivered through a website and be based on at least 2 of the 6 core ACT processes: acceptance, contact with the present moment, cognitive defusion, self-as-context, values, and committed action. This criterion was set in accordance with Swain et al's [14] reasoning that setting a minimum of 2 ACT processes will include interventions that address multiple aspects of the ACT model (rather than just a stand-alone technique), while simultaneously avoiding the reiteration of earlier review research on single processes (eg, pure mindfulness). Furthermore, studies emphasizing stress reduction (rather than anxiety reduction) or cognitive restructuring techniques were excluded in the present review [14]. Only articles written in the English language and published in peer-reviewed journals were included.

Search Strategy

The following electronic databases of scientific research were searched: EMBASE, MEDLINE, ProQuest Central, PsycINFO,

Scopus, and Web of Science. Keywords used to search each database were ((Anxiety) AND ((ACT) OR (acceptance and commitment therapy) OR (iACT)) AND ((Internet) OR (online) OR (web-based) OR (etherapy))). The search was limited to peer-reviewed papers published in English up to the month of September 2018. The reference lists of included papers were manually searched. The Association for Contextual Behavioral Science (ACBS) website [35] was also surveyed for iACT research. The ACBS organization is a worldwide online learning and research community for ACT.

Article Selection

Upon removal of duplicate citations, article titles and abstracts were then independently screened and appraised next to the eligibility criteria by 2 researchers (JK and AR). Full text review was again conducted by both reviewers, and divergent views were resolved through discussion and mutual agreement.

Data Extraction

Data from included studies were extracted by one reviewer (JK) into a standardized coding sheet and then checked by a second reviewer (AC). Data items extracted for synthesis included the following:

1. Reference source: first author surname and year of publication.
2. Anxiety-related problem/disorder under investigation.
3. Study design: methodology, comparator trial arms, and measurement points.
4. Population: country, participants, sample size, age, and gender breakdown.
5. iACT intervention details: intervention name, manual protocol, number of modules, treatment length, therapist guidance, and educational content.
6. Efficacy: study quality rating, anxiety measures, outcomes, and effect sizes.
7. Adherence: attrition rate, treatment satisfaction, and intention-to-treat analysis.

Attrition in this review was defined and measured as the relative number of participants who commenced but did not complete the active iACT treatment phase.

Quality Assessment

Öst's [36] *psychotherapy outcome study methodology rating form* (POMRF) was used to appraise the quality of the reviewed iACT articles. It is a 22-item questionnaire that addresses key methodological issues pertinent to psychotherapy intervention studies such as research design, participant details, psychotherapy implementation, therapist characteristics, user adherence, statistical analyses, and clinical significance [36]. It can be used to assess study quality across multiple research designs, including RCTs, and has been used in an earlier systematic review by Swain et al [14] on ACT for anxiety. Each item in the POMRF is rated on a 3-point scale ranging from 0 (*poor*) to 2 (*good*), with the total POMRF score calculated as the sum of all points. The total score for any given study can range from 0 to 44 points, with higher scores indicating greater overall methodological quality. Öst [36] found the POMRF questionnaire to demonstrate high internal consistency (Cronbach alpha of 0.86) and interrater reliability (kappa coefficient mean of 0.75). In this systematic review, the included iACT studies were independently assessed next to the POMRF items by 2 reviewers (JK and AC) who individually calculated the overall POMRF scores. On completion, these scores were compared between reviewers, and the discrepancies were resolved by discussion and mutual agreement.

Narrative Synthesis

A narrative synthesis approach was chosen for this systematic review due to a paucity of iACT papers and the heterogeneity found between studies (eg, populations, designs, comparators, and anxiety outcomes). Results are presented as descriptive data with no further analyses performed.

Results

Study Selection

The literature search yielded 566 papers. A total of 186 records remained after duplicate removal. Of these records, 20 met the eligibility criteria (see Figure 1).

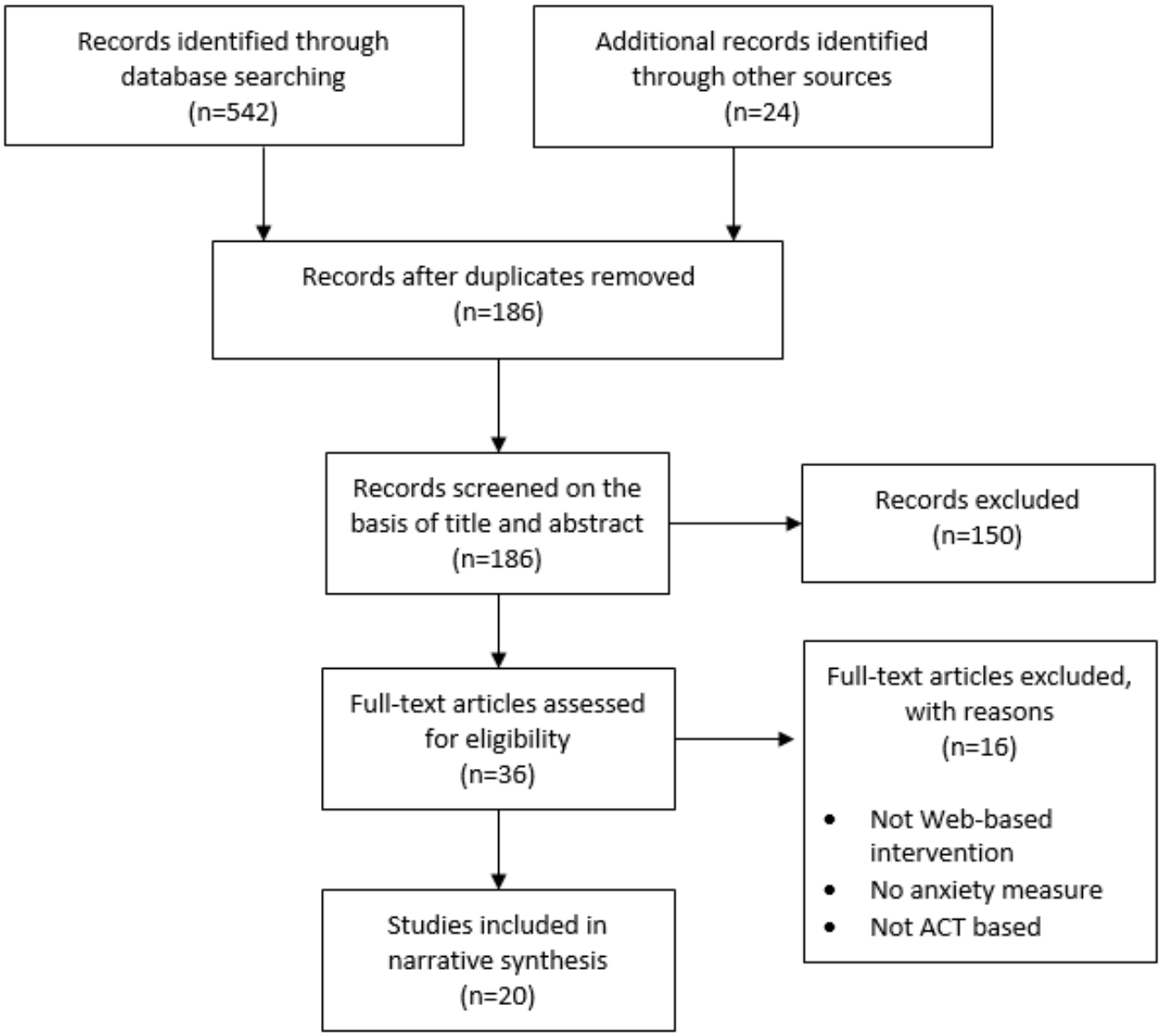
Participant Characteristics

Table 1 shows that the majority of studies and their participants were from Sweden (35%, 7/20) and the United States (35%, 7/20), with the remaining studies from Australia (10%, 2/20), the Netherlands (10%, 2/20), Finland (5%, 1/20), and Denmark (5%, 1/20). All participants were adults with mean ages ranging from 18.37 to 59 years. Sample sizes ranged from 13 to 238 participants, with a median of 76 participants. Participants were mainly female with an average sample proportion of 72.62% (range 43.4%-100%). Both clinical (65%, 13/20) and nonclinical (35%, 7/20) studies were conducted. Anxiety symptom reduction was typically measured as a secondary outcome with 15 out of 20 studies (75%) targeting people with a general health issue and/or comorbid anxiety problem. Several studies (25%, 5/20) targeted a specific anxiety disorder. Diagnoses of GAD and SAD were obtained by clinical interviews delivered in person or by phone, whereas IAD diagnoses were obtained by videoconferencing.

Research Designs

Table 2 shows that just over half of the studies were RCTs (55%, 11/20). Control groups consisted of wait-lists (n=8), online discussion forums (n=2), expressive writing (n=2), iCBT program (n=1), and a mental health education website (n=1). The remaining studies in the review were uncontrolled pilot studies (45%, 9/20). There were 12 studies that reported between-group analyses, and 14 that reported within-group analyses, on their respective anxiety measures. Pretest, posttest, and follow-up assessments were used in 14 out of 20 studies (70%). The follow-up assessments ranged from 2 weeks to 1 year, with a median of 3 months.

Figure 1. Flowchart of systematic review search results. ACT: Acceptance and Commitment Therapy.



Internet-Delivered Acceptance and Commitment Therapy Interventions

Educational content was typically presented in multimedia formats (ie, audio, video, text, images, and animations) with interactive exercises, worksheets, quizzes, and homework assignments provided (see Table 3). Several studies (40%, 8/20) based their content directly upon a detailed treatment manual that is publicly available. Every study appraised the iACT treatment of individuals, as no group-based approaches were adopted. Treatment lengths ranged from 2 to 12 weeks, with a mode of 8. The number of treatment modules ranged from 2 to 9, with a mode of 7. Most studies (70%, 14/20) tested iACT with therapist guidance, except for several (30%, 6/20) which delivered iACT as pure self-help.

Attrition and Treatment Satisfaction

Attrition in the active iACT treatment phase varied between the studies, with a mean of 19.19% and a range of 0 to 60.77% (see Table 4). To handle missing data, 18 out of 20 studies reported the use of an intention-to-treat analysis (90%). Several studies (55%, 9/20) used a standardized measure to collect treatment satisfaction ratings from participants after iACT completion. These measures included the Client Satisfaction Questionnaire-8 items (CSQ-8) and System Usability Scale (SUS). Several studies (35%, 7/20) incorporated nonstandardized questions (NSQ) on treatment satisfaction. Participants' mean rating of their iACT experience ranged from above average to high satisfaction.

Anxiety Measures

Details of the standardized anxiety measures used are summarized in Table 5. The anxiety subscales of the Depression Anxiety Stress Scales-21 items and Hospital Anxiety and Depression Scale were the most frequently used measures (7 and 5 studies, respectively). Most studies (70%, 14/20) used a single anxiety measure, whereas the remaining 6 studies (30%) used multiple measures.

Anxiety Outcomes

Nearly all studies (90%, 18/20) reported small-to-large improvement on anxiety symptoms after treatment (see Table 5). Two studies on GAD found large anxiety reductions immediately after iACT, and at the 3- [40] and 6-month [39]

follow-up points. Two studies on SAD found large reductions in anxiety symptoms after 8 weeks of iACT and at 3-month follow-up [42,43]. Hoffmann et al [45] found large effect sizes on both anxiety measures at posttest and 3-month follow-up for IAD. Anxiety symptoms also reduced for people with fibromyalgia [51], tinnitus [44], chronic pain [37,55], major depression [38], a history of interpersonal trauma [41], and mild to moderate symptoms of depression, anxiety, and stress [46-48,50,53,54,56].

However, the study by Levin et al [49] found no significant subclinical anxiety reduction after 3 weeks of self-help iACT treatment. Similarly, Räsänen et al [54] detected no significant difference between an iACT and a wait-list control group on anxiety symptoms among university students. Furthermore, Molander et al [52] found no significant difference in anxiety between their iACT and wait-list control group at post-treatment among adults with hearing problems. When comparing iACT to iCBT on anxiety reduction in tinnitus sufferers, Hesser et al [44] found no significant differences. No significant differences were found between iACT with or without therapist support for people with SAD [43]. Finally, the study by Trompetter et al [55] found no significant difference between iACT and expressive writing for reducing anxiety symptoms in chronic pain sufferers.

Quality Assessment

Total POMRF scores for each study are presented in Table 5. Out of a possible total of 44 points, the mean score for the reviewed iACT studies was 19 (SD 5.7), with a range of 10 to 29. The included studies gave either fair (65%, 13/20) or good (35%, 7/20) descriptions on their research participants in terms of inclusion/exclusion criteria, demographics, and comorbidity.

Table 1. Description of participants in reviewed internet-delivered Acceptance and Commitment Therapy (iACT) studies.

Source	Country	Problem/diagnosis	Clinical	N[a]	Age (years), mean (SD)	Females in total sample, n (%)
Buhrman [37]	Sweden	Chronic pain	Yes	76	49.1 (10.34)	45 (59)
Carlbring [38]	Sweden	Major depression	Yes	80	44.4 (13.5)	66 (82)
Dahlin [39]	Sweden	GAD[b]	Yes	103	39.48 (10.73)	86 (83.5)
Dahlin [40]	Sweden	GAD	Yes	14	32 (10)	11 (78)
Fiorillo [41]	United States	Interpersonal trauma	Yes	25	39.12 (16)	25 (100)
Gershkovich [42]	United States	SAD[c]	Yes	13	33.2 (10.04)	9 (69)
Gershkovich [43]	United States	SAD	Yes	42	31.5 (9.95)	27 (64)
Hesser [44]	Sweden	Tinnitus	Yes	99	48.5 (14.7)	43 (43)
Hoffmann [45]	Denmark	IAD[d]	Yes	15	38.8 (NR[e])	12 (67)
Kelson [46]	Australia	DAS[f] symptoms	No	40	21.62 (2.4)	21 (52)
Levin [47]	United States	DAS symptoms	No	76	18.37 (.54)	41 (54)
Levin [48]	United States	DAS symptoms	No	82	21.88 (3.50)	62 (75)
Levin [49]	United States	DAS symptoms	No	234	21.61 (5.48)	180 (77)
Levin [50]	United States	DAS symptoms	No	79	20.51 (2.73)	52 (66)
Ljótsson [51]	Sweden	Fibromyalgia	Yes	41	52 (9)	41 (100)
Molander [52]	Sweden	Hearing problems	Yes	61	59 (11)	41 (67.2)
Pots [53]	The Netherlands	Depressive symptoms	Yes	236	46.85 (12.06)	179 (75.8)
Räsänen [54]	Finland	DAS symptoms	No	68	24.29 (3.28)	58 (85.3)
Trompetter [55]	The Netherlands	Chronic pain	Yes	238	52.8 (12.37)	181 (76)
Viskovich [56]	Australia	DAS symptoms	No	130	26.3 (7.96)	98 (75)

[a]Refers to total number of participants.

[b]GAD: generalized anxiety disorder.

[c]SAD: social anxiety disorder.

[d]IAD: Illness anxiety disorder.

[e]NR: not reported.

[f]DAS: depression, anxiety, and stress.

Table 2. Description of research designs in reviewed internet-delivered Acceptance and Commitment Therapy (iACT) studies.

Source	RCT[a]	Trial arms (n)[b]	Design	Measurements
Buhrman [37]	Yes	iACT (38); forum (38)	BG[c]	Pre[d], Post[e], 6-month FU[f]
Carlbring [38]	Yes	iACT (40); WLC[g] (40)	BG	Pre, Post, 3-month FU
Dahlin [39]	Yes	iACT (52); WLC (51)	WG[h]; BG	Pre, Post, 6-month FU
Dahlin [40]	No	iACT (14)	WG	Pre, Post, 2 to 3-month FU
Fiorillo [41]	No	iACT (25)	WG	Pre, Post
Gershkovich [42]	No	iACT (13)	WG	Pre, Mid[i], Post, 3-month FU
Gershkovich [43]	No	iACT (22); iACT + TS[j] (20)	WG, BG	Pre, Mid, Post
Hesser [44]	Yes	iACT (35); iCBT[k] (32); forum (32)	BG	Pre, Post, 1-year FU
Hoffmann [45]	No	iACT (15)	WG	Pre, Post, 3-month FU
Kelson [46]	No	iACT (40)	WG	Pre, Post, 2-week FU
Levin [47]	Yes	iACT (37); WLC (39)	WG; BG	Pre, Post, 3-week FU
Levin [48]	No	iACT (82)	WG	Pre, Post
Levin [49]	Yes	iACT (114); website (120)	WG; BG	Pre, Post, 3-month FU
Levin [50]	Yes	iACT (40); WLC (39)	WG; BG	Pre, Post
Ljótsson [51]	No	iACT (41)	WG	Pre, Post, 6-month FU
Molander [52]	Yes	iACT (31); WLC (30)	BG	Pre, Post
Pots [53]	Yes	iACT (82); EW[l] (67); WLC (87)	BG	Pre, Post, 3 and 6-month FU
Räsänen [54]	Yes	iACT (33); WLC (35)	WG; BG	Pre, Post, 1-year FU
Trompetter [55]	Yes	iACT (82); EW (79); WLC (77)	BG	Pre, Post, 3-month FU
Viskovich [56]	No	iACT (130)	WG	Pre, Post

[a]RCT: randomized controlled trial.

[b]Refers to number of participants.

[c]BG: between-groups.

[d]Pre: pretest.

[e]Post: posttest.

[f]FU: follow-up.

[g]WLC: wait-list control.

[h]WG: within-groups.

[i]Mid: midtest.

[j]TS: therapist support.

[k]iCBT: internet-delivered cognitive behavioral therapy.

[l]EW: expressive writing.

All studies' participants were either somewhat or very representative of people seeking treatment for their respective anxiety-related issue. Diagnostic reliability was obtained in all 13 clinical studies with a structured interview from a trained assessor. Several studies (40%, 8/20) recruited participants with anxiety conditions inclusive of both acute chronicity (ie, less than one year duration) and low severity. Of the remaining studies, 5 consisted of chronic anxiety sufferers with at least moderate severity [37,41,45,51,55], and the other 7 included people with subthreshold symptoms [46-48,50,54,56].

Most studies (70%, 14/20) were rated poorly on research design because of being uncontrolled or only making use of a wait-list control group (see Table 2). The control for concomitant treatments outside of iACT was also generally poor (65%, 13/20). Several studies (45%, 9/20) reported a power analysis. Just over half of the studies (60%, 12/20) randomly assigned participants to treatment conditions, with 6 of these also randomly assigning therapists [38,39,44,53-55]. Assessment points were typically conducted at pretest, posttest, and follow-up of less than a year (60%, 12/20). Two studies had follow-up of a year or more [44,54]. All iACT studies employed adequate statistical analyses and presented results fully with means and SDs.

All studies incorporated moderate or very specific anxiety outcome measures with good psychometric properties for their targeted conditions. Details on the clinical significance of

therapy outcomes were provided in several studies (45%, 9/20), 5 of which used Jacobson and Truax's [57] criteria for clinically significant change [39,41,43,44,53]. There was no blinding of assessors across any included iACT studies. Outcome measures were collected via the internet in the form of self-report questionnaires.

About 8 of the iACT interventions under investigation were based on manual treatment programs (see Table 3). All therapist-guided iACT studies (70%, 14/20) deployed at least 2 therapists. Graduate psychology students were often used to guide participants through the iACT interventions (55%, 11/20). Therapist competence was checked either somewhat or frequently in all guided iACT studies. The equality of therapy hours between conditions was good in the 3 studies that had an active treatment control [44,53,55]. Treatment adherence checks were made in all studies and proportions of attrition rates were provided. Intent-to-treat analyses were performed in all but 2 studies (see Table 4).

Table 3. Details of the internet-delivered Acceptance and Commitment Therapy (iACT) interventions studied.

Source	Name	Manual	Modules	Length (weeks)	Guided	System features
Buhrman [37]	Not specified	No	7	7	Yes	Audio, text, interactive exercises
Carlbring [38]	The Depression Help	No	7	8	Yes	Multimedia[a], animations, interactive exercises, homework, CD
Dahlin [39]	The Worry Help	No	7	9	Yes	Multimedia, animations, interactive exercises, homework, CD
Dahlin [40]	The Worry Help	No	7	8-10	Yes	Multimedia, animations, interactive exercises, homework, CD
Fiorillo [41]	Not specified	Yes	6	6	No	Multimedia, interactive exercises, worksheets
Gershkovich [42]	Not specified	Yes	8	8	Yes	Multimedia, interactive exercises, homework, quizzes
Gershkovich [43]	Not specified	Yes	8	8	Yes	Multimedia, interactive exercises, homework, quizzes
Hesser [44]	Not specified	Yes	8	8	Yes	Multimedia, interactive exercises, homework
Hoffmann [45]	Not specified	Yes	7	12	Yes	Multimedia, interactive exercises, homework
Kelson [46]	FearLess	No	9	2	No	Text, images, interactive exercises, quizzes
Levin [47]	ACT on College Life	No	2	3	No	Multimedia, animations, interactive exercises, worksheets, quizzes
Levin [48]	ACT on College Life	No	3	4	Yes	Multimedia, animations, interactive exercises, worksheets, quizzes
Levin [49]	ACT on College Life	No	2	3	No	Multimedia, animations, interactive exercises, worksheets, quizzes
Levin [50]	Not specified	No	6	4	No	Multimedia, interactive exercises, homework, worksheets,
Ljótsson [51]	Not specified	Yes	5	10	Yes	Text, interactive exercises, worksheets
Molander [52]	Not specified	No	8	8	Yes	Text, images, audio, interactive exercises, homework
Pots [53]	Not specified	Yes	9	12	Yes	Multimedia, interactive exercises, diary
Räsänen [54]	The Student Compass	No	9	7	Yes	Multimedia, interactive exercises, journal
Trompetter [55]	Living with Pain	Yes	9	12	Yes	Text, interactive exercises, diary
Viskovich [56]	You Only Live Once	No	4	4	No	Multimedia, interactive exercises, animations

[a]Multimedia: combined use of audio, video, text, and images.

Table 4. Details of attrition rates and treatment satisfaction.

Source	Attrition (%)	ITT[a]	Satisfaction measures	Satisfaction outcomes
Buhrman [37]	17.14	Yes	—[b]	—
Carlbring [38]	0	Yes	—	—
Dahlin [39]	19.2	Yes	NSQ[c]	High satisfaction
Dahlin [40]	0	Yes	CSQ-8[d]	High satisfaction
Fiorillo [41]	16	Yes	CSQ-8	High satisfaction
Fiorillo [41]	16	Yes	SUS[e]	High satisfaction
Gershkovich [42]	0	Yes	NSQ	High satisfaction
Gershkovich [43]	31	Yes	NSQ	High satisfaction
Hesser [44]	8.57	Yes	—	—
Hoffmann [45]	20	No	—	—
Kelson [46]	26	No	SUS	Above average satisfaction
Levin [47]	8	Yes	SUS	High satisfaction
Levin [48]	56	Yes	SUS	High satisfaction
Levin [48]	56	Yes	NSQ	High satisfaction
Levin [49]	30	Yes	SUS	Above average satisfaction
Levin [50]	20	Yes	SUS	Above average
Levin [50]	20	Yes	NSQ	Adequate satisfaction
Ljtósson [51]	2	Yes	—	—
Molander [52]	16.13	Yes	—	—
Pots [53]	15	Yes	—	—
Räsänen [54]	12	Yes	NSQ	High satisfaction
Trompetter [55]	28.05	Yes	CSQ-8	High satisfaction
Trompetter [55]	28.05	Yes	NSQ	More than adequate satisfaction
Viskovich [56]	60.77	Yes	SUS	Above average satisfaction

[a]ITT: intention-to-treat analysis.

[b]not applicable.

[c]NSQ: nonstandardized questions.

[d]CSQ-8: Client Satisfaction Questionnaire-8 items.

[e]SUS: System Usability Scale.

Table 5. Details on study quality, anxiety measures, outcomes, and effect sizes (Cohen *d*) for internet-delivered Acceptance and Commitment Therapy (iACT).

Source	POMRF[a]	Measures	WG[b] effect size	BG[c] effect size
Buhrman [37]	23	HADS-A[d]	—[e]	0.18 (iACT vs WLC[f])
Carlbring [38]	22	BAI[g]	—	0.45 (iACT vs WLC)
Dahlin [39]	23	PSWQ[h]	1.35 (pre[i] to post[j])	0.87 (iACT vs WLC)
Dahlin [39]	23	GAD-7[k]	1.89 (pre to post)	0.98 (iACT vs WLC)
Dahlin [39]	23	GAD-Q-IV[l]	1.5 (pre to post)	0.70 (iACT vs WLC)
Dahlin [39]	23	BAI	0.92 (pre to post)	0.55 (iACT vs WLC)
Dahlin [39]	20	PSWQ	2.14 (pre to post)	—
Fiorillo [41]	20	DASS-A[m]	0.89 (pre to post)	—
Gershkovich [42]	18	SPAI-SP[n]	1.47 (pre to post)	—
Gershkovich [42]	18	LSAS[o]	0.92 (pre to post)	—
Gershkovich [43]	22	SPAI-SP	1.07 (pre to post)	NS[p]
Gershkovich [43]	22	LSAS	0.77 (pre to post)	NS
Gershkovich [44]	22	SIAS[q]	1.01 (pre to post)	NS
Hesser [44]	29	HADS-A	—	0.59 (iACT vs Forum)
Hoffmann [45]	22	WI-7[r]	1.63 (pre to post)[s]	—
Hoffmann [45]	22	SCL-92 Anx[t]	0.73 (pre to post)[s]	—
Kelson [46]	11	DASS-A	0.42 (pre to FU[u])	—
Kelson [46]	11	GAD-7	0.66 (pre to post)	—
Levin [47]	13	DASS-A	0.95 (pre to FU)	0.81 (iACT vs WLC)
Levin [48]	14	DASS-A	0.55 (pre to post)	—
Levin [49]	11	DASS-A	NS	NS
Levin [50]	13	CCAPS-Anx[v]	0.39 (pre to post)	NS
Levin [50]	13	CCAPS-Social	0.69 (pre to post)	0.78 (iACT vs WLC)
Ljótsson [51]	18	HADS-A	0.75 (pre to post)	—
Molander [52]	17	GAD-7	—	NS
Pots [53]	27	HADS-A	—	0.49 (iACT vs WLC)
Pots [53]	27	HADS-A	—	0.41 (iACT vs EW[w])
Räsänen [54]	19	DASS-A	0.42 (pre to post)	*Ns* (iACT vs WLC)
Trompetter [55]	28	HADS-A	—	0.39 (iACT vs WLC)
Trompetter [55]	28	HADS-A	—	*Ns* (iACT vs EW)
Viskovich [56]	10	DASS-A	0.32 (pre to post)	—

[a]POMRF: psychotherapy outcome study methodology rating form.

[b]WG: within-group.

[c]BG: between-group.

[d]HADS-A: Hospital Anxiety and Depression Scale-Anxiety subscale.

[e]N/A: not applicable.

[f]WLC: wait-list control.

[g]BAI: Beck Anxiety Inventory.

[h]PSWQ: Penn State Worry Questionnaire.

[i]pre: pretest.

[j]post: posttest.
[k]GAD-7: Generalized Anxiety Disorder-7 item.
[l]GAD-Q-IV Generalized Anxiety Disorder Questionnaire IV.
[m]DASS-A: Depression, Anxiety and Stress Scales-Anxiety subscale.
[n]SPAI-SP: Social Phobia and Anxiety Inventory-Social Phobia subscale.
[o]LSAS: Liebowitz Social Anxiety Scale-Total.
[p]NS: nonsignificant.
[q]SIAS: Social Interaction Anxiety Scale.
[r]WI-7: Whiteley Index-7.
[s]Standardized response mean effect size.
[t]SCL-92: Symptom Checklist-92.
[u]FU: follow-up.
[v]CCAPS: Counseling Center Assessment of Psychological Symptoms.
[w]EW: expressive writing.

Discussion

Overview

In total, 18 of the 20 studies found in this systematic review reported small-to-large anxiety reductions among participants after iACT on standardized measures. Studies typically aimed to treat a primary issue other than an anxiety disorder. Just over half were RCTs, with a fewer number of randomized or uncontrolled pilot studies of new iACT interventions. The interventions were typically rich in multimedia information content and interactive exercises that were provided to participants with therapist guidance over a period of several weeks. Attrition rates were commensurate to other website interventions based on other therapy models [32], and treatment satisfaction was rated above average to high among participants who completed their respective iACT program. There is still considerable need to test iACT on all anxiety conditions with studies of greater methodological quality to corroborate extant findings.

Participant Characteristics

The delivery of iACT for anxiety is a new field as all studies identified in this review were published from 2012 onward. Many studies were pioneering new iACT programs for people in their respective countries, with Sweden and the United States featured prominently. There were few to no identified studies from other major Western countries such as the United Kingdom, Australia, New Zealand, and Canada. Participants in the included iACT studies were primarily young to middle-aged adults with a high representation of females, indicating that there is substantial need to improve the generalizability of results by creating or applying existing iACT interventions to the wider population across other demographics (eg, children, adolescents, the elderly, and male participants).

Research Designs

Nine of the iACT studies were uncontrolled pilot studies. Preliminary intervention testing with open and uncontrolled research designs has merit for cost-effectively recruiting participants and demonstrating feasibility of the program on key issues of usability, safety, and limited efficacy [34]. Future creators of iACT programs can also glean insight from these studies on how to design, develop, and test their new interventions for phase I clinical trials. However, uncontrolled pilot studies do not allow for solid conclusions on the efficacy of treating anxiety with iACT because of the lack of control for confounding factors, such as concomitant treatments. For instance, the effect sizes of the uncontrolled studies may be inflated as they are susceptible to threats of internal validity such as maturation, history, and regression to the mean [17]. Therefore, the magnitude of the reported effect sizes in this review cannot be compared across uncontrolled and controlled studies. The remaining studies had stronger research designs that used wait-list control and active treatment groups for comparators. Overall, more RCTs are required to corroborate the promising findings of the uncontrolled pilot studies.

Internet-Delivered Acceptance and Commitment Therapy Interventions

Presently, there is no single or leading iACT intervention for anxiety. Multiple programs have been developed and tested with key differences. For instance, some iACT programs were tailored with information and exercises to people from specific groups such as university students [47], or chronic pain sufferers [55]. Few iACT interventions had more than one study investigating their therapeutic impact. Due to the heterogeneity between programs, it is not clear what modules were well received by participants and facilitated therapeutic change. Future corelational and component analytical studies on which modules/exercises are better suited for certain populations could help identify optimal features and predictors of success between interventions [13].

Despite the differences between iACT interventions, there were still commonalities. This included the multimedia representation of content, proposed mechanisms of change underlying exercises (eg, acceptance and mindfulness), number of intervention modules, and overall treatment length of several weeks. Another key shared factor was therapist guidance, with most studies having used therapists to varying extents. Research has reported that therapist-guided interventions, in general, produce larger effects than unguided interventions [20]. However, Gershkovich et al [43] found equal anxiety reductions delivering iACT with or without therapist support. Furthermore, RCT studies by Levin et al [47,50] found significant reductions in subclinical anxiety symptoms with unguided iACT. These outcomes should be examined in further research. If the findings hold across studies

with greater methodological rigor, then unguided iACT could potentially help address therapist worker shortages [22,23].

Attrition and Treatment Satisfaction

Attrition rates in the active treatment phase of iACT varied considerably between studies. Three studies in particular had a high percentage of participants not completing the entire program [38,48,56]. Carlbring et al [38] used no treatment satisfaction measures for their iACT program, so it was not possible to gauge the extent to which this impacted upon adherence. In contrast, Levin et al [48] had high system usability and treatment satisfaction ratings from program completers despite a high attrition rate. Similar positive ratings were also received across all other studies that included treatment satisfaction, measures regardless of attrition rate. This suggests that iACT may not be suitable for everyone but can be particularly well received by some as an online treatment modality. The mean attrition rate across all included studies (19.19%) was lower than the median dropout rate of 24% identified across other Web-based interventions, and lower than the weighted average of 35% [32]. Future research would need to make equivalency comparisons on attrition rates between iACT programs and mainstream iCBT programs to help determine if iACT has superior user retention.

Anxiety Outcomes

Generalized Anxiety Disorder

Two studies indicated that iACT can be efficacious and acceptable for adults with GAD [39,40]. The first study by Dahlin et al [40] found a large reduction on GAD symptomatology across 14 participants. However, this uncontrolled pilot study had a small sample, which limits conclusions on the true efficacy of *The Worry Help* program because of possible treatment confounds. However, large reductions across multiple measures of GAD were also found in a subsequent RCT study by Dahlin et al [39]. There was no clinically significant deterioration throughout this study, with GAD symptoms improving beyond the acute treatment phase [40]. A key limitation of this study, however, was the use of only a wait-list control rather than an active treatment group. Nevertheless, both of these studies still provide early support for treating adults with GAD using therapist-guided iACT over a period ranging from 8 to 10 weeks. This pattern of results mirrors the large GAD anxiety reductions of face-to-face ACT treatment [12], which implicates the potential of guided iACT to provide adjunctive clinical care.

Social Anxiety Disorder

Two studies indicate that iACT can be efficacious and acceptable for adults with SAD [42,43]. The first study by Gershkovich et al [42] found large effects on multiple measures of SAD symptoms across 13 participants. However, this uncontrolled pilot study had a small sample, which limits conclusions on the true efficacy of this iACT program because of possible treatment confounds. The other study by Gershkovich et al [43] investigated the same iACT program with participants randomized to 1 of 2 conditions: treatment with or without therapist support. Once more there were large reductions in SAD anxiety symptoms found in both groups at posttest with 14 out of 26 (53.8%) participants no longer meeting SAD criteria. Key limitations of this study were the lack of a non-iACT control group, and no participant follow-up assessments. These foundational studies provide tentative support for treating adults with SAD using iACT over an 8-week period. This pattern of results mirrors the large SAD anxiety reductions found with face-to-face ACT treatment [12], which further implicates the potential of iACT to provide adjunctive clinical care.

Illness Anxiety Disorder

It appears that iACT holds promise as an efficacious treatment for IAD, a condition of severe health anxiety that was previously known as hypochondriasis before the DSM-5 [45]. Significantly large reductions on both measures of anxiety were found at post-treatment that were also sustained at 3-month follow-up. This implies that therapist-guided iACT over 12 weeks is a feasible treatment approach for severe health anxiety among adults. However, this study was an uncontrolled pilot study with a small sample size of 15 self-referred patients. Therefore, these preliminary results must be validated in further larger-scale RCT research to control for treatment confounds.

General Anxiety Symptoms

Small-to-large improvements in general anxiety symptoms were reported across most studies that evaluated iACT on various physical and mental health issues. These outcomes were found across several different Western countries, which indicate the capacity for iACT to transcend national and cultural borders. Furthermore, the highest quality RCT evidence in this review revealed that iACT was just as efficacious as iCBT [44], and expressive writing [53,55]. These studies had the highest POMRF scores in the review with a range of 27 to 29. These scores are all well above the average POMRF score of 18.1 reported earlier for face-to-face ACT treatment RCTs [36]. Additionally, iACT outcomes were maintained at follow-up for all studies that found significant reductions. The primary implication is that iACT can provide a transdiagnostic, efficacious, and acceptable treatment option for people with general anxiety symptoms.

However, the results of studies that investigated the benefit of iACT for general anxiety symptoms, specifically among young adult university students, were mixed. Several studies indicated that iACT can be efficacious for reducing anxiety symptoms in this population with moderate-to-large effects [46-48,50]. Results in 2 of these studies [47,48] were procured from online treatment using fewer ACT-based modules (2 to 3) than the other iACT studies (5 to 9). This suggests that it could be possible to streamline iACT content to produce equivalent results to programs with more content. Furthermore, 2 of these studies [46,47] had no therapist guidance, and the pilot study by Kelson et al [46] found significant subclinical anxiety reduction after 2 weeks. These findings hold promise for delivering unguided iACT interventions in the form of self-help, or stepped care, to the "worried well" to procure rapid anxiety relief [6].

However, while the RCT by Räsänen et al [54] found a moderate within-group effect of guided iACT treatment on anxiety

reduction in young adults, there was no significant difference between the iACT condition and the wait-list control group at posttest. Furthermore, an RCT by Levin et al [49] did not find a significant anxiety reduction after delivering 2 modules of unguided iACT treatment over a 3-week period. This outcome contradicts the significant results of the earlier RCT study by Levin et al [47] that delivered the same iACT intervention without therapist guidance over 3 weeks. Together, these results imply that iACT may not be efficacious for all young adults, or that possible confounds occurred, such as wait-list participants using concomitant treatments, or natural fluctuations in subclinical symptomatology.

It is important to note that all of the unguided iACT studies that included participants with subclinical anxiety had low POMRF ratings. Therefore, methodological issues limit conclusive interpretations of these early stage outcomes. For instance, the first study by Levin et al [47] was a small feasibility trial that used a less conservative alpha level to determine significance (P<.10). The studies by Kelson et al [46] and Levin et al [48] made use of single-group, uncontrolled open-trial designs. The third study by Levin et al [49] had low power and did not use a dosage equivalent treatment condition to help control for nonspecific confounds. Pilot study recruitment issues of university students may also have affected results. For instance, most participants in the most recent study by Levin et al [50] reportedly sought the incentive of psychology course credits more so than mental health treatment which could have impacted upon observed program engagement, satisfaction, and self-report outcomes. Therefore, subsequent studies would need to appraise the impact of iACT in university settings among a pure treatment seeking a sample that does not receive any inducements such as money or course credits. Despite these limitations, there is still a need for more iACT research on people suffering from subclinical general anxiety symptoms, as they can be quite debilitating [5]. Therefore, more methodologically rigorous RCT studies are needed to clarify the utility of unguided iACT for subclinical anxiety.

Future Research

Presently, there is substantial need to test iACT interventions on all anxiety-related problems. There were no iACT studies on separation anxiety disorder, selective mutism, specific phobia, panic disorder, agoraphobia, substance/medication-induced anxiety disorder, or anxiety disorders due to medical conditions [2]. Questions also remain regarding the capacity of iACT interventions to enact therapeutic change in vein of established treatments, such as iCBT, on multiple factors (eg, cost-effectiveness, credibility, brevity, outcome longevity, and integration with clinical services). Therefore, further investigation into the impact of individual ACT-based modules, with or without therapist guidance, using RCT designs that incorporate an active positive control is warranted. This could help pinpoint optimal treatment time lengths, clarify the quality of content, and elucidate the utility of unguided treatment protocols. Future research should also consider quality appraisal guidelines such as Öst's [36] POMRF criteria when designing iACT studies. This would help address gaps in the current literature, such as concomitant treatment confounds, lack of clinical significance reporting, and anxiety outcome sustainability beyond a year. As most of the iACT programs were researched by the people who developed them, independent investigation would clarify the validity of extant findings.

Review Limitations

This systematic review has several limitations. Owing to the inclusion of only peer-reviewed English-language journal articles, there may potentially be more anxiety iACT studies in the public domain available in other languages or formats (eg, conference proceedings). Even among the English language–based literature reviewed, there were studies found which met inclusion criteria, yet were not self-described as iACT. For example, several studies used terms such as "acceptance-based" or "acceptance and values-based behavior therapy" in their article titles [38-40,51]. It is possible that a study may have been overlooked because of obscure nomenclature. Also, this systematic review only focused on Web-based interventions and did not include in its scope other online modalities, such as mobile phone apps, social media, and virtual reality systems. Furthermore, there were no subgroup analyses performed on the impact of iACT on anxiety. Thus, insight was lacking into the applicability of iACT to people with specific demographic characteristics (eg, gender, age, or education level).

With regard to quality appraisal limitations, all types of studies were included in this review to reduce publication bias. Although non-RCT studies can provide insight into iACT on anxiety, there is a greater risk of bias that needs to be accounted for when interpreting the validity of the reported effects. This risk of bias was addressed with Öst's [36] POMRF questionnaire in this review, yet there is still room for subjectively interpreting POMRF criteria. For example, a study may have actually met certain criteria in its execution, yet scored lower on an item in our quality appraisal because of the authors omitting that information in the publication text. Although the subjective interpretation of risk of bias is an issue with any quality appraisal method, it is still a limitation to consider that may impact on the capacity to replicate the findings of this systematic review.

Conclusions

Nearly all studies reported the beneficial impact of iACT on anxiety. Small to large reductions in anxiety symptoms among populations suffering from GAD, SAD, IAD, and anxiety-related health problems were found in all but 2 studies. More research is required to establish outcomes on iACT for other anxiety relevant conditions such as panic disorder, agoraphobia, separation anxiety, selective mutism, and specific phobias [2]. An important aspect of this is making direct comparisons of iACT treatments with established active interventions (eg, iCBT) on key standardized measures of anxiety, adherence, usability, and treatment satisfaction. Further research on iACT for anxiety on untested demographics (eg, children, adolescents, and the elderly), as well as in countries and cultural settings outside of Sweden and the United States, is also warranted. Overall, the current findings indicate that iACT can be an effective and acceptable treatment for some anxiety conditions among young to middle aged adults in Western societies.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist.

[PDF File (Adobe PDF File), 67KB - jmir_v21i1e12530_app1.pdf]

References

1. Baxter AJ, Scott KM, Vos T, Whiteford HA. Global prevalence of anxiety disorders: a systematic review and meta-regression. Psychol Med 2013 May;43(5):897-910. [doi: 10.1017/S003329171200147X] [Medline: 22781489]
2. Diagnostic and statistical manual of mental disorders (DSM-5). Washington DC: American Psychiatric Association; 2013.
3. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
4. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen H. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
5. Haller H, Cramer H, Lauche R, Gass F, Dobos GJ. The prevalence and burden of subthreshold generalized anxiety disorder: a systematic review. BMC Psychiatry 2014 May 01;14:128 [FREE Full text] [doi: 10.1186/1471-244X-14-128] [Medline: 24886240]
6. Wick JY, Zanni GR. Hypochondria: the worried well. Consult Pharm 2008 Mar;23(3):192-208. [doi: 10.4140/TCP.n.2008.192] [Medline: 18454583]
7. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
8. Bomyea J, Lang AJ, Craske MG, Chavira D, Sherbourne CD, Rose RD, et al. Suicidal ideation and risk factors in primary care patients with anxiety disorders. Psychiatry Res 2013 Aug 30;209(1):60-65 [FREE Full text] [doi: 10.1016/j.psychres.2013.03.017] [Medline: 23608160]
9. Eifert GH, Forsyth JP. Acceptance and Commitment Therapy for Anxiety Disorders: a Practitioner's Treatment Guide to Using Mindfulness, Acceptance, and Values-Based Behavior Change. Oakland, CA: New Harbinger Publications; 2005.
10. Hayes SC, Levin ME, Plumb-Vilardaga J, Villatte JL, Pistorello J. Acceptance and commitment therapy and contextual behavioral science: examining the progress of a distinctive model of behavioral and cognitive therapy. Behav Ther 2013 Jun;44(2):180-198 [FREE Full text] [doi: 10.1016/j.beth.2009.08.002] [Medline: 23611068]
11. Bluett EJ, Homan KJ, Morrison KL, Levin ME, Twohig MP. Acceptance and commitment therapy for anxiety and OCD spectrum disorders: an empirical review. J Anxiety Disord 2014 Aug;28(6):612-624. [doi: 10.1016/j.janxdis.2014.06.008] [Medline: 25041735]
12. Dalrymple KL, Herbert JD. Acceptance and commitment therapy for generalized social anxiety disorder: a pilot study. Behav Modif 2007 Sep;31(5):543-568. [doi: 10.1177/0145445507302037] [Medline: 17699117]
13. Roemer L, Orsillo SM, Salters-Pedneault K. Efficacy of an acceptance-based behavior therapy for generalized anxiety disorder: evaluation in a randomized controlled trial. J Consult Clin Psychol 2008 Dec;76(6):1083-1089 [FREE Full text] [doi: 10.1037/a0012720] [Medline: 19045976]
14. Ruiz F. A review of Acceptance and Commitment Therapy (ACT) empirical evidence: Correlational, experimental psychopathology, component and outcome studies. Rev Int Psicol Ter Psicol 2010;10(1):125-162 [FREE Full text]
15. Swain J, Hancock K, Hainsworth C, Bowman J. Acceptance and commitment therapy in the treatment of anxiety: a systematic review. Clin Psychol Rev 2013 Dec;33(8):965-978. [doi: 10.1016/j.cpr.2013.07.002] [Medline: 23999201]
16. Beck AT, Dozois DJ. Cognitive therapy: current status and future directions. Annu Rev Med 2011 Feb;62:397-409. [doi: 10.1146/annurev-med-052209-100032] [Medline: 20690827]
17. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 2006 Jan;26(1):17-31. [doi: 10.1016/j.cpr.2005.07.003] [Medline: 16199119]
18. Koen N, Stein DJ. Pharmacotherapy of anxiety disorders: a critical review. Dialogues Clin Neurosci 2011;13(4):423-437 [FREE Full text] [Medline: 22275848]
19. Hayes SC. Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. Behav Ther 2004;35(4):639-665. [doi: 10.1016/S0005-7894(04)80013-3]
20. 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]

21. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. Ann Behav Med 2009 Aug;38(1):4-17. [doi: 10.1007/s12160-009-9130-7] [Medline: 19787305]
22. 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 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
23. Gamm L, Stone S, Pittman S. Rural Health Research. College Station, TX: The Texas A&M University System Health Science Center, School of Rural Public Health, Southwest Rural Health Research Center; 2010. Rural healthy people 2010: a companion document to healthy people 2010 URL: https://www.ruralhealthresearch.org/projects/286 [accessed 2019-01-15] [WebCite Cache ID 75RvN58qI]
24. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
25. 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]
26. Andersson G, Carlbring P, Ljótsson B, Hedman E. Guided internet-based CBT for common mental disorders. J Contemp Psychother 2013 Dec;43(4):223-233. [doi: 10.1007/s10879-013-9237-9]
27. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
28. Grist R, Cavanagh K. Computerised cognitive behavioural therapy for common mental health disorders, what works, for whom under what circumstances? A systematic review and meta-analysis. J Contemp Psychother 2013 Dec;43(4):243-251. [doi: 10.1007/s10879-013-9243-y]
29. Newby JM, Twomey C, Yuan LS, Andrews G. Transdiagnostic computerised cognitive behavioural therapy for depression and anxiety: a systematic review and meta-analysis. J Affect Disord 2016 Jul 15;199:30-41. [doi: 10.1016/j.jad.2016.03.018] [Medline: 27060430]
30. Zettle RD. The evolution of a contextual approach to therapy: from comprehensive distancing to ACT. Int J Behav Consult Ther 2005;1(2):77-89. [doi: 10.1037/h0100736]
31. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression: systematic review. J Med Internet Res 2009 Apr;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
32. Melville KM, Casey LM, Kavanagh DJ. Dropout from internet-based treatment for psychological disorders. Br J Clin Psychol 2010 Nov;49(4):455-471. [doi: 10.1348/014466509X472138] [Medline: 19799804]
33. Brown M, Glendenning A, Hoon AE, John A. Effectiveness of web-delivered acceptance and commitment therapy in relation to mental health and well-being: a systematic review and meta-analysis. J Med Internet Res 2016 Aug 24;18(8):e221 [FREE Full text] [doi: 10.2196/jmir.6200] [Medline: 27558740]
34. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
35. Association for Contextual Behavioral Science. URL: https://contextualscience.org/ [accessed 2019-01-16] [WebCite Cache ID 75SrINDQ2]
36. Öst LG. Efficacy of the third wave of behavioral therapies: a systematic review and meta-analysis. Behav Res Ther 2008 Mar;46(3):296-321. [doi: 10.1016/j.brat.2007.12.005] [Medline: 18258216]
37. Buhrman M, Skoglund A, Husell J, Bergström K, Gordh T, Hursti T, et al. Guided internet-delivered acceptance and commitment therapy for chronic pain patients: a randomized controlled trial. Behav Res Ther 2013 Jun;51(6):307-315. [doi: 10.1016/j.brat.2013.02.010] [Medline: 23548250]
38. Carlbring P, Hägglund M, Luthström A, Dahlin M, Kadowaki Å, Vernmark K, et al. Internet-based behavioral activation and acceptance-based treatment for depression: a randomized controlled trial. J Affect Disord 2013 Jun;148(2-3):331-337. [doi: 10.1016/j.jad.2012.12.020] [Medline: 23357657]
39. Dahlin M, Andersson G, Magnusson K, Johansson T, Sjögren J, Håkansson A, et al. Internet-delivered acceptance-based behaviour therapy for generalized anxiety disorder: a randomized controlled trial. Behav Res Ther 2016 Feb;77:86-95. [doi: 10.1016/j.brat.2015.12.007] [Medline: 26731173]
40. Dahlin M, Ryberg M, Vernmark K, Annas N, Carlbring P, Andersson G. Internet-delivered acceptance-based behavior therapy for generalized anxiety disorder: a pilot study. Internet Interv 2016 Nov;6:16-21 [FREE Full text] [doi: 10.1016/j.invent.2016.08.004] [Medline: 30135810]
41. Fiorillo D, McLean C, Pistorello J, Hayes SC, Follette VM. Evaluation of a web-based acceptance and commitment therapy program for women with trauma-related problems: a pilot study. J Contextual Behav Sci 2017 Jan;6(1):104-113. [doi: 10.1016/j.jcbs.2016.11.003]
42. Gershkovich M, Herbert JD, Forman EM, Glassman L. Guided internet-based self-help intervention for social anxiety disorder with videoconferenced therapist support. Cogn Behav Pract 2016 May;23(2):239-255 [FREE Full text] [doi: 10.1016/j.cbpra.2015.08.005]

43. Gershkovich M, Herbert JD, Forman EM, Schumacher LM, Fischer LE. Internet-delivered acceptance-based cognitive-behavioral intervention for social anxiety disorder with and without therapist support: a randomized trial. Behav Modif 2017 Sep;41(5):583-608. [doi: 10.1177/0145445517694457] [Medline: 28776431]
44. Hesser H, Gustafsson T, Lundén C, Henrikson O, Fattahi K, Johnsson E, et al. A randomized controlled trial of internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. J Consult Clin Psychol 2012 Aug;80(4):649-661. [doi: 10.1037/a0027021] [Medline: 22250855]
45. Hoffmann D, Rask CU, Hedman-Lagerlöf E, Ljótsson B, Frostholm L. Development and feasibility testing of internet-delivered acceptance and commitment therapy for severe health anxiety: pilot study. JMIR Ment Health 2018 Apr 6;5(2):e28 [FREE Full text] [doi: 10.2196/mental.9198] [Medline: 29625957]
46. Kelson JN, Lam MK, Keep M, Campbell AJ. Development and evaluation of an online acceptance and commitment therapy program for anxiety: phase I iterative design. J Technol Hum Serv 2017 Apr;35(2):135-151. [doi: 10.1080/15228835.2017.1309311]
47. Levin ME, Pistorello J, Seeley JR, Hayes SC. Feasibility of a prototype web-based acceptance and commitment therapy prevention program for college students. J Am Coll Health 2014 Jan;62(1):20-30 [FREE Full text] [doi: 10.1080/07448481.2013.843533] [Medline: 24313693]
48. Levin ME, Pistorello J, Hayes SC, Seeley JR, Levin C. Feasibility of an acceptance and commitment therapy adjunctive web-based program for counseling centers. J Couns Psychol 2015 Jul;62(3):529-536. [doi: 10.1037/cou0000083] [Medline: 25961755]
49. Levin ME, Hayes SC, Pistorello J, Seeley JR. Web-based self-help for preventing mental health problems in universities: comparing acceptance and commitment training to mental health education. J Clin Psychol 2016 Mar;72(3):207-225. [doi: 10.1002/jclp.22254] [Medline: 26784010]
50. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-based acceptance and commitment therapy for mental health problems in college students: a randomized controlled trial. Behav Modif 2017 Jan 1;41(1):141-162. [doi: 10.1177/0145445516659645] [Medline: 27440189]
51. Ljótsson B, Atterlöf E, Lagerlöf M, Andersson E, Jernelöv S, Hedman E, et al. Internet-delivered acceptance and values-based exposure treatment for fibromyalgia: a pilot study. Cogn Behav Ther 2014;43(2):93-104. [doi: 10.1080/16506073.2013.846401] [Medline: 24215278]
52. Molander P, Hesser H, Weineland S, Bergwall K, Buck S, Jäder Malmlöf J, et al. Internet-based acceptance and commitment therapy for psychological distress experienced by people with hearing problems: a pilot randomized controlled trial. Cogn Behav Ther 2018 Mar;47(2):169-184. [doi: 10.1080/16506073.2017.1365929] [Medline: 28895787]
53. Pots WT, Fledderus M, Meulenbeek PA, ten Klooster PM, Schreurs KM, Bohlmeijer ET. Acceptance and commitment therapy as a web-based intervention for depressive symptoms: randomised controlled trial. Br J Psychiatry 2016 Jan;208(1):69-77. [doi: 10.1192/bjp.bp.114.146068] [Medline: 26250745]
54. Räsänen P, Lappalainen P, Muotka J, Tolvanen A, Lappalainen R. An online guided ACT intervention for enhancing the psychological wellbeing of university students: a randomized controlled clinical trial. Behav Res Ther 2016 Mar;78:30-42. [doi: 10.1016/j.brat.2016.01.001] [Medline: 26848517]
55. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KM. Internet-based guided self-help intervention for chronic pain based on acceptance and commitment therapy: a randomized controlled trial. J Behav Med 2015 Feb;38(1):66-80. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]
56. Viskovich S, Pakenham KI. Pilot evaluation of a web-based acceptance and commitment therapy program to promote mental health skills in university students. J Clin Psychol 2018 Jul;74(12):2047-2069. [doi: 10.1002/jclp.22656] [Medline: 29962090]
57. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037/0022-006X.59.1.12] [Medline: 2002127]

Abbreviations

ACBS: Association for Contextual Behavioral Science
ACT: Acceptance and Commitment Therapy
CBT: cognitive behavioral therapy
CSQ-8: Client Satisfaction Questionnaire-8 Items
DSM-5: Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition
GAD: generalized anxiety disorder
iACT: internet-delivered Acceptance and Commitment Therapy
IAD: illness anxiety disorder
iCBT: internet-delivered Cognitive-Behavioral Therapy
NSQ: nonstandardized questions
POMRF: psychotherapy outcome study methodology rating form
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCT: randomized controlled trial
SAD: social anxiety disorder
SUS: System Usability Scale

Edited by G Eysenbach; submitted 18.10.18; peer-reviewed by SC Hayes, D Hoffmann; comments to author 01.11.18; revised version received 10.11.18; accepted 24.11.18; published 29.01.19.

Please cite as:
Kelson J, Rollin A, Ridout B, Campbell A
Internet-Delivered Acceptance and Commitment Therapy for Anxiety Treatment: Systematic Review
J Med Internet Res 2019;21(1):e12530
URL: http://www.jmir.org/2019/1/e12530/
doi:10.2196/12530
PMID:30694201

Original Paper

Electronic Health Program to Empower Patients in Returning to Normal Activities After Colorectal Surgical Procedures: Mixed-Methods Process Evaluation Alongside a Randomized Controlled Trial

Chantal M den Bakker[1,2], MD; Judith AF Huirne[3], MD, PhD; Frederieke G Schaafsma[1], MD, PhD; Charlotte de Geus[1], BSc; Hendrik J Bonjer[2], MD, PhD; Johannes R Anema[1], MD, PhD

[1]Department of Occupational and Public Health, Amsterdam Public Health Research Institute, VU University Medical Center, Amsterdam, Netherlands
[2]Department of Surgery, VU University Medical Center, Amsterdam, Netherlands
[3]Department of Gynecology, VU University Medical Center, Amsterdam, Netherlands

Corresponding Author:
Chantal M den Bakker, MD
Department of Occupational and Public Health
Amsterdam Public Health Research Institute
VU University Medical Center
Van der Boechorststraat 7
Amsterdam, 1081 BT
Netherlands
Phone: 31 20 4445703
Fax: 31 20 4448387
Email: c.denbakker@vumc.nl

Abstract

Background: Long-term recovery takes longer than expected despite improved surgical techniques and Enhanced Recovery After Surgery programs. An electronic health (eHealth) care program ("ikherstel") was developed to partially substitute perioperative care for patients undergoing colorectal surgical procedures. Successfully tested eHealth programs are not always implemented in usual care, and it is, therefore, important to evaluate the process to optimize future implementation.

Objective: The aim of this study was to evaluate whether the eHealth intervention was executed as planned.

Methods: A mixed-methods process evaluation was carried out alongside a multicenter randomized controlled trial (RCT). This evaluation was performed using the Linnan and Steckler framework for the quantitative part of this study, measuring the components reach, dose delivered, dose received, fidelity, and participants' attitudes. Total implementation scores were calculated using the averaging approach, in which the sum of all data points is divided by the number of data points and the total adherence to the protocol is measured. For the qualitative part, the Unified Theory of Acceptance and Use of Technology framework was used. The quantitative data were based on participants' questionnaires, a logistic database, a weblog, and participants' medical files and were obtained by performing semistructured interviews with participants of the RCT.

Results: A total of 151 participants of 340 eligible patients were included in the RCT, of which 73 participants were allocated to the intervention group. On the basis of the quantitative process data, total implementation scores for the website, mobile app, electronic consult, and activity tracker were 64%, 63%, 44%, and 67%, respectively. Participants in the qualitative part experienced the program as supportive and provided guidance on their recovery process after colorectal surgery. Most frequently mentioned barriers were the limited interaction with and feedback from health care professionals and the lack of tailoring of the convalescence plan in case of a different course of recovery.

Conclusions: The intervention needs more interaction with and feedback from health care professionals and needs more tailored guidance in case of different recovery or treatment courses. To ensure a successful implementation of the program in daily practice, some adjustments are required to optimize the program in a blended care form.

Trial Registration: Netherlands Trial Registry NTR5686; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=5686 (Archieved by WebCite at http://www.webcitation.org/75LrJaHrr)

(J Med Internet Res 2019;21(1):e10674) doi:10.2196/10674

KEYWORDS

return to normal activities; return to work; patient reported outcome measures; colectomy; process evaluation

Introduction

Background

In-hospital stay after colorectal surgical procedures has shortened enormously attributable to improved surgical techniques and Enhanced Recovery After Surgery programs, but further recovery at home still takes longer than expected by health care professionals [1-4]. Electronic health (eHealth) can be a suitable tool to optimize perioperative care including at-home recovery by providing tailored information, increasing patients' self-management, and by delivering interactive communication features. Patients can become their own empowered and motivated health managers [5-8].

To partially substitute guiding and monitoring of long-term recovery including resumption of normal activities and work of colorectal patients, an eHealth intervention called "ikherstel" or "I recover" was developed using the intervention mapping protocol. This innovative eHealth program consists of, among others, a website, a mobile app, an activity tracker, and the possibility of an electronic consult (eConsult). The program will be evaluated in a multicenter single blinded randomized controlled trial (RCT) [9].

Many eHealth care programs are developed and tested in health care; however, successful eHealth programs will not automatically be implemented in usual daily care in all cases. An integral part of evaluating successful eHealth interventions is measuring the adherence to the intervention protocol, as this will play an important role in interpreting the results regarding the effectivity and it might improve further implementation [10-12]. Measuring adherence and compliance to perioperative care processes is a fundamental aspect in improving the quality of surgical care [13]. It is also desirable for optimal implementation to evaluate how well the intervention was appreciated by participants, and a process evaluation can contribute to this [14]. The quantitative data in a mixed-methods approach contribute to understanding why a (complex) intervention has its intended impact, if any, and in which domain this went as planned or not [15]. In addition, by using qualitative data, patients' experiences including barriers and facilitators may be reviewed in more detail to adjust the eHealth care program for future implementation.

Objectives

The aim of this study was to evaluate whether the recovery-orientated eHealth intervention was executed as planned. This can help to conduct per protocol analyses, to assist with interpreting the future trial outcomes, and to determine important factors for program scale up in this specific cancer-dominated study population.

Methods

Trial Design

A mixed-methods process evaluation of quantitative data obtained from participants' questionnaires, a logistic database, a weblog, and participants' medical files and a qualitative analysis of semiconstructed interviews were conducted. This evaluation was carried out alongside a multicenter, single-blinded RCT in 10 teaching hospitals in the Netherlands and is reported in accordance with the Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and onLine TeleHealth (CONSORT-EHEALTH) [16]. The intervention development including trial protocol has been published previously [17]. This RCT was approved by the Medical Ethics Committee of the VU University medical center under registration number 2014.301. Under registration number NTR5686, this study was also registered at the Netherlands Trial Registry.

Participants

Patients aged between 18 and 75 years who were scheduled for a laparoscopic or abdominal colorectal resection or hysterectomy were eligible to participate in the RCT. This process evaluation describes all participants who underwent a colorectal resection with malignant and benign indications. Exclusion criteria were (1) surgery without a curative intention, (2) concomitant surgical procedures, (3) not able to use the internet, (4) unable to understand Dutch questionnaires, and (5) receiving neoadjuvant treatment. For the interviews, purposeful sampling was used including participants with both a positive and a negative rating of the program.

Intervention

Participants were randomized and allocated to the intervention or control group in a 1:1 ratio by a researcher who was independent from the recruitment, data collection process, or analyses. Study participants were blinded to the allocation. Participants in the control group received usual care and access to a placebo website, which contained a patient information brochure about the surgical procedure. Participants in the intervention group received access to an innovative eHealth care program (called "ikherstel"-intervention or "I recover"-intervention). This program consisted, among others, of a website, a mobile phone app, an activity tracker, and the possibility to ask questions to health care professionals of their own hospital via an eConsult. All included functionalities can be found in the intervention mapping study [17].

Data Collection

Participants filled out an adherence and satisfaction questionnaire 3 months after the surgical procedure. In addition, a logistic database, a weblog, and participants' medical files were used to collect quantitative data. Semistructured interviews were conducted in October and November 2017, which was

approximately 1 year after inclusion of the first patient in the RCT who participated in the qualitative part of this study. The topics were created to gather information about patients' barriers and facilitators for use of the intervention. In the preparation phase, topics and questions were created based on literature, the theoretical framework, and discussion with the project team. All interviews were recorded with a voice recorder. Informed consent was obtained initially when patients participated in the RCT.

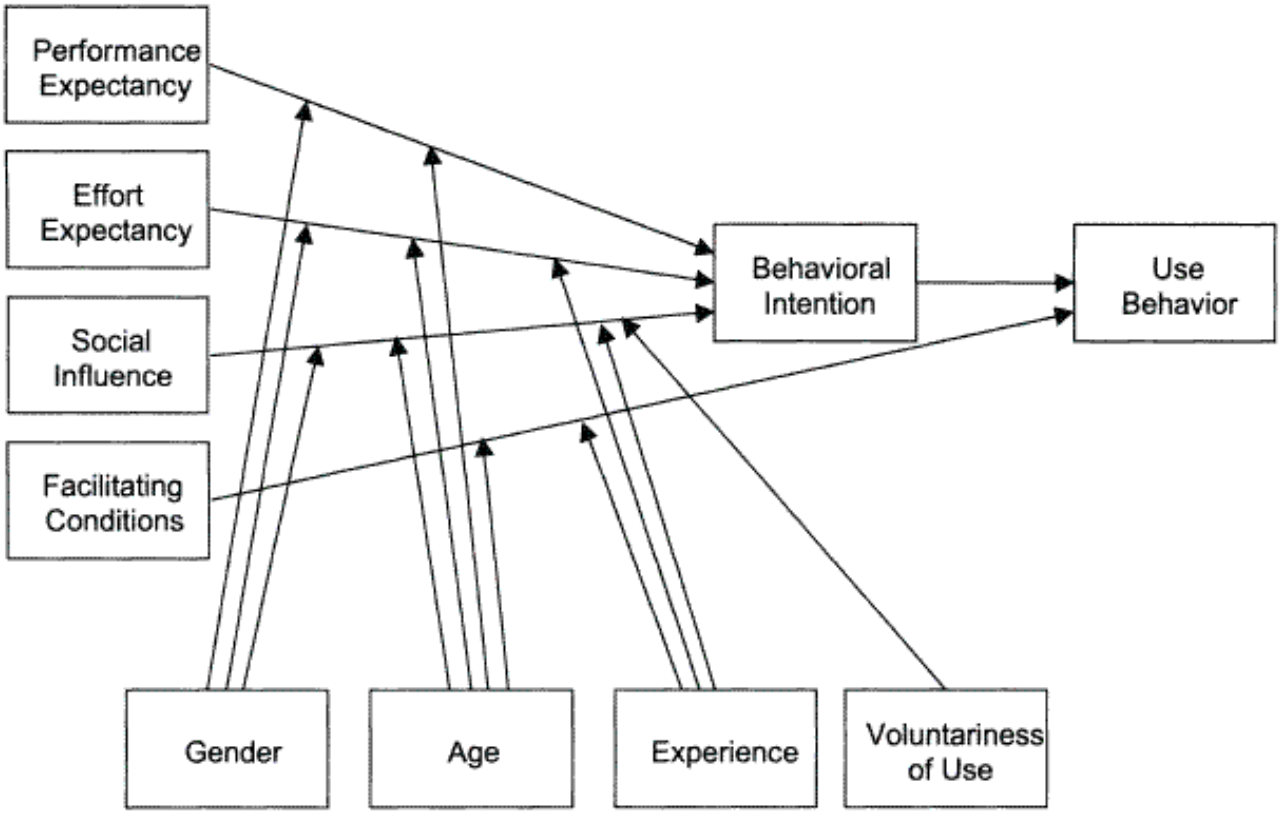
Theoretical Frameworks and Process Outcomes

The Linnan and Steckler model was used for the quantitative part of the process evaluation [18]. This is a commonly used model and has the potential to systematically evaluate the process of implementation. The adherence to the intervention is described in 5 terms: (1) the proportion of intended target audience that participated in the study—reach, (2) the number of intended units of each component that was delivered to the intervention group—dose delivered, (3) the number of participants from the intervention group that actively engaged the delivered components of the intervention—dose received, (4) the extent of the intervention that was delivered as planned—fidelity, and (5) participants' satisfaction and usage barriers of the intervention—participants' attitudes. The components were assessed for each function separately, except the component participants' attitudes for eConsult. Operationalization per component per process outcome is further explained in the Results section. Total implementation scores were calculated by using the averaging approach, in which the sum of all data points is divided by the number of data points and the total adherence to the protocol is measured. The Unified Theory of Acceptance and Use of Technology framework (UTAUT) was used for the qualitative part of this evaluation [19]. This framework integrates fragmented theories and research on individual acceptance of information technology into 1 model with 4 core constructs to help gain insight into (1) the degree patients' believe that using the "ikherstel" program will help in their recovery—performance expectancy, (2) the ease of use patients experienced when using the "ikherstel" program—effort expectancy, (3) the degree to which patients perceive that it is important others believe that he or she should use the "ikherstel" program—social influence, and (4) other external factors that facilitate or inhibit the use of the "ikherstel" program—facilitating conditions. These constructs measure the impact of behavioral intention and use behavior. These can play a role as determinants of user acceptance. Overall, 4 moderators (gender, age, experience, and voluntariness of use) were incorporated into this framework. However, it was decided to not measure these moderators, given the qualitative nature of this part of the study. A detailed description of this framework is provided in Figure 1.

Data Analysis

IBM SPSS Statistics version 22.0 was used for analyzing the quantitative data. These data were analyzed using descriptive statistics such as frequencies, means, and SDs. The semistructured interviews for the qualitative part were recorded and transcribed verbatim. Each participant was allocated a study number, and all names were removed from the transcripts to ensure an anonymous analysis. After transcription, the researchers first familiarized themselves with the data and read the transcripts thoroughly. Subsequently, the verbatim transcripts were analyzed comprising open, axial, and selective coding.

Figure 1. The Unified Theory of Acceptance and Use of Technology framework (UTAUT) for the qualitative part of the study.



Overall, 25% of the transcripts were coded and analyzed by a second independent researcher to reduce investigator bias and improve validity and inter-rater reliability. Furthermore, the level of data saturation was systematically studied. The frequency of quotes within each theme and their distribution across the interviews were explored, based on a data saturation approach as described by Guest et al [20]. Cited quotes were translated directly from Dutch and were added to illustrate the themes. The qualitative data analysis ATLAS.ti software (version 7.0, Scientific Software Development) was used.

Results

Quantitative Part

Reach

During February 2016 and September 2017, a total of 426 patients scheduled for a colorectal resection were invited to participate in the RCT. Of these patients, 62 patients were ultimately not eligible to participate and 24 patients did not reach on time. Of the 340 suitable patients, 151 patients agreed to participate in the trial (44.4%) and gave informed consent. A total of 73 colorectal participants were allocated to the intervention group and received the recovery-orientated eHealth program. The flow diagram of the inclusion process is presented in Figure 2. Moreover, 69% (50/73) of the patients were male and their mean age was 62.6 years, and 56 patients (77%, 56/73) had a colon or rectal carcinoma. All baseline characteristics are presented in Table 1. A total of 61 participants completed the quantitative evaluation questionnaire 3 months after surgery, which is considered a representative sample with respect to all baseline characteristics. All other scores per domain per intervention functionality and how this has been operationalized in this study are presented in Table 2.

Figure 2. Flow diagram of the inclusion process.

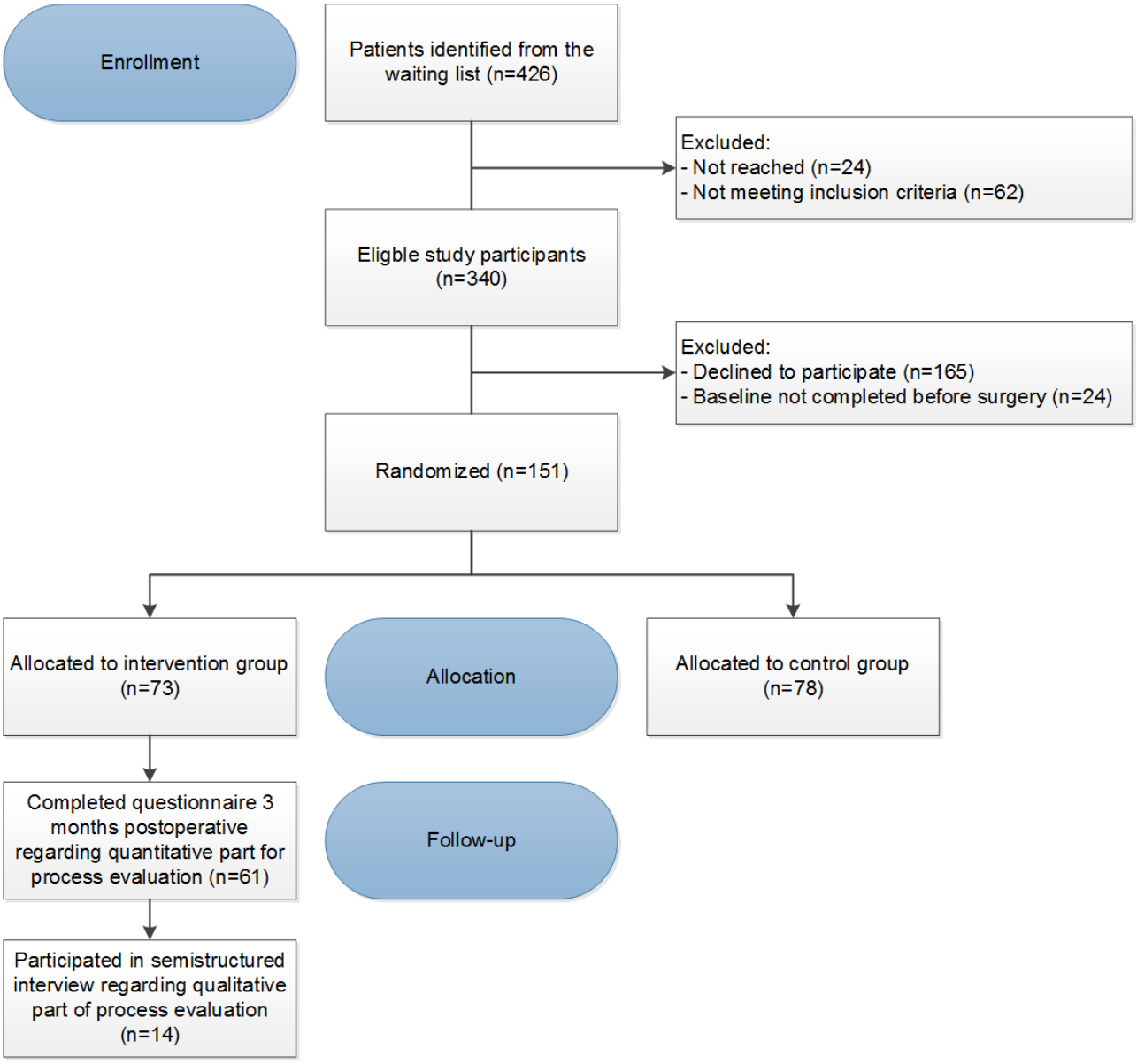


Table 1. Baseline characteristics (N=73).

Variable	Intervention group
Gender, n (%)	
Female	23 (32)
Male	50 (69)
Age, mean (SD)	62.6 (7.8)
Nationality, n (%)	
Dutch	73 (100)
Other	0 (0)
Education level, n (%)	
Low	20 (27)
Medium	31 (43)
High	22 (30)
Work status, n (%)	
Employed	36 (49)
Not employed	37 (51)
Type of surgery, n (%)	
Abdominal procedure	4 (6)
Laparoscopic procedure	69 (95)
Indication, n (%)	
Benign	17 (23)
Malignant	56 (77)

Table 2. Results of the quantitative part of the study according to Linnan and Steckler. 151 (44.4%) patients who met the inclusion criteria signed informed consent and were randomized to the intervention or control group.

Reach	Website	App	Electronic consult	Activity tracker
Dose delivered	73 (100%, 73/73) patients who received an account/all patients of the intervention group	59 (81%, 59/73) patients who received an account for the app/all patients of the intervention group	73 (100%, 73/73) patients who received an account/all patients of the intervention group	57 (78%, 57/73) patients who received an activity tracker/all patients of the intervention group
Data collection	Logistic database	Logistic database	Logistic database	Logistic database
Dose received	60 (82%, 60/73) patients who made a convalescence plan/patients who received an account	39 (66%, 39/59) patients who installed the app/patients who received an account and completed the questionnaire	4 (6%, 4/73) patients who asked a question about the Web portal/patients who received an account for the Web portal	39 (68%, 39/57) patients who connected the activity tracker to their phone/all patients that received an activity tracker
Data collection	Weblog	Weblog	Weblog	Weblog
Fidelity	17 (28%, 17/60) patients who used the website/patients who made a convalescence plan	24 (62%, 24/39) patients who used the app/patients who installed the app and completed the questionnaire	1 (25%, 1/4) questions that are answered/questions that are asked	30 (77%, 30/39) patients that used the activity tracker/all patients that connected an activity tracker
Data collection	Questionnaire	Questionnaire	Weblog	Weblog+Questionnaire
Participants' attitude	Mean score 7.1 (1-9)	Mean score 7.5 (1-10)	N/A[a]	Mean score 7.1 (1-10)
Data collection	Questionnaire	Questionnaire	N/A	Questionnaire
Implementation score	64% (the sum of all data points/by the number of data points)	63% (the sum of all data points/by the number of data points)	44% (the sum of all data points/by the number of data points)	67% (the sum of all data points/by the number of data points)

[a]N/A: not applicable.

Dose Delivered

All participants (100%) received an account for the intervention website and thereby the ability to use the eConsult functionality (100%). Overall, 59 participants had a suitable mobile phone for the mobile app (81%, 59/73), and 57 participants had a suitable smartphone to use the activity tracker (78%, 57/73).

Dose Received

Overall, 60 participants made a convalescence plan on the intervention website (82%, 60/73). A total of 39 participants downloaded the app (66%, 39/59) and connected the activity tracker with their smartphone (68%, 39/57). Moreover, 4 participants used the possibility of an eConsult (6%, 4/73).

Fidelity

Overall, 17 participants used the website more than once (28%). Of the 39 participants who had downloaded the app, 24 participants used it (62%). In addition, 13 participants used the app in combination with the website and the remaining 11 participants only used the app. Overall, 4 participants only used the website. Moreover, 1 out of the 4 questions asked via the eConsult option was answered by a health care professional (25%). Of the 39 participants that connected the activity tracker, 30 participants actually used it (77%).

Participants' Attitude

Mean ratings for the website, app, and activity tracker were 7.1 (range 1-9), 7.5 (range 1-10), and 7.1 (range 1-10), respectively.

Implementation Scores

Total implementation scores for the website, the mobile app, the eConsult, and the activity tracker are 64%, 63%, 44%, and 67%, respectively.

Qualitative Part

A total of 14 semistructured interviews were conducted during November and December 2017. The overall mean age at the time of the interview was 62 years (range 45-76 years); 4 participants were female and 10 were male. The characteristics of these 14 participants are shown in Multimedia Appendix 1. Most important findings are discussed below per construct regarding the UTAUT model. An overview of all findings is presented in Table 3.

Performance Expectancy

The key themes (1) guidance and support, (2) information provision, (3) communication, and (4) functionality were identified out of the analysis.

Guidance and Support

Participants were positive about the "ikherstel" program. They found it supporting and guiding in their recovery process after colorectal surgery. Participants stated that the program was a good guideline and a good way of monitoring their progress. This knowledge gave the participants a feeling of reassurance and security about their recovery:

> *Security. A feeling of confidence, whether I am going in the right direction. Naturally, you do not continuously call a doctor or the hospital to check whether it is going okay, or I feel this or I feel that. Because of this app you know, you have to meet these requirements, so it is all right.* [Participant 3, male, 54 years]

The participants also perceived the program as effective in progressing their recovery, whereby participants felt that the "ikherstel" program had achieved its goal. Furthermore, the participants were positive about the activity tracker or the concept of an activity tracker. Participants agreed with the notion that using the activity tracker motivated them to be active and that it was a good way to reflect on their level of activity. It provided a goal to work toward and the steps they had to take to get to this goal. Participants stated that the difference between what they thought their level of activity was and what the activity tracker showed could either motivate them to be more active the following day or positively surprised them and generated a positive and satisfied feeling:

> *But that does stimulate you at the end of the day, to see where I am and "oh tomorrow I have to do a bit more."* [Participant 12, female, 65 years]

When participants did not fully adhere to the personalized recovery plan, this was generally because participants recovered faster than the recovery plan advised to them. Therefore, the provided recommendations were sometimes reviewed as too conservative. As a result, these participants would follow their own instinct and resumed activities when their body felt ready for it.

Information Provision

Participants stated that the information provided on the website was insightful, useful, and all-encompassing. However, not all patients felt the need to read the extra information on the website as they already received sufficient information and guidance from their treating health care professionals. Participants mentioned 3 advantages of the provided information on the website: (1) the information was always available; (2) the information came from a reputable and, therefore, trustworthy source; and (3) the information was more elaborate than they had received in the hospital. These advantages were considered relevant as participants mentioned that it was easy to forget what was told in the hospital and they did not always have time to discuss everything in the hospital:

> *It is also information that you otherwise do not get. Visiting the doctor is always pretty quick of course. The time I was in the hospital, yes you can of course ask all the questions to the nurse, but this was just a little bit more. And you can also look it up again. It was nice.* [Participant 12, female, 65 years]

Table 3. Key findings for each of the 4 constructs of the Unified Theory of Acceptance and Use of Technology model.

Construct	Topic	Elaboration
Performance expectancy	• Supporting and guiding	• The "ikherstel" program supported and guided the recovery process after surgery. It provides good opportunities to monitor your own recovery process as well as a goal to work toward.
	• Activity tracker motivates	• The activity tracker motivated to be physically active.
	• Deviated from recovery plan	• Some participants did not adhere to the personalized recovery plan. Reasons for deviation included advanced recovery before the recovery plan resulting in resumed activities when the participants' body felt ready for it.
	• Psychological aspects should be included	• Participants stated there is too much focus on physical recovery, whereas psychological aspects were not taken into account. • According to the participants, psychological well-being and a positive attitude to cope with the emotional burden of cancer diagnosis were an integral part of recovery.
	• More personalized	• Participants would have preferred it to be more personalized, including more focus on individual aspects and needs of the patient and inclusion of social conditions that influence the recovery process.
	• Useful and insightful	• The provided information on the website was found to be useful and all-encompassing. • Some participants indicated that they did not need the information or that they did not read it. • Some participants also desired extra information on diet and prevention and more extensive information on symptoms and complications.
	• Information on the websites provides advantages	• It was easy to forget what was said in the hospital and the website provided a good backup. • Participants only had a short amount of time to ask questions in the hospital, and therefore, it was good to have the website that provided additional information. • The information on the website was readily available. • Participants felt the information was trustworthy as it came from a reputable information source.
	• Need for more feedback and interaction	• There was a need for more personal interaction and feedback on progress of the recovery process. • There was a one-way information stream from the patient to the "ikherstel" program; this has to become a two-way information stream.
	• More involvement of hospital or doctor	• Participants desired more involvement and feedback from the hospital or treating doctor.
	• Functionalities have to work correctly	• The activity tracker experienced problems with connection and did not function properly in some cases. • After consultation via electronic consult, no answer was given, whereas this should have been given within 2 days.
Effort expectancy	• Easy to use	• The "ikherstel" program was found to be easy to use and it costed no effort.
	• More support in setting up the program	• Some participants stated they would have appreciated more support during the program's start-up phase.
Social influence	• Would recommend the program	• The majority of participants would recommend the "ikherstel" program to their family and friends.
	• Social influence from health professional	• Family and friends had little or no social influence. • There was more social influence from the hospital and doctor.
Facilitating and inhibiting conditions	• Inflexible in case of alternative disease course	• An inhibiting factor was the inflexibility of the personalized recovery plan in case of a deviant recovery course, which can occur in case of complications.

Construct	Topic	Elaboration
	• Insufficient for prolonged disease course (chemotherapy)	• The "ikherstel" program was insufficient for patients who received chemotherapy due to the prolonged disease course and the additional needs (eg, additional information on helpful methods to cope with chemotherapy and prolonged use activity tracker).
	• Level of information provision from hospital	• The information provision from the hospital influenced the need for the "ikherstel" program. This could be either a negative influence in case of adequate information provision or positive in case of a lack of information provision.
	• Positive attitude	• A positive attitude facilitated the recovery process and therefore also the use of the "ikherstel" program.
	• Physical fitness before surgery	• Physical fitness before surgery influenced the postoperative recovery process.

Communication

Despite the advantages of the "ikherstel" program, participants stated that there is still much to be improved in relation to communication. Currently, there is mostly a one-way information stream from the user to the "ikherstel" program with the exception of eConsult. Participants would also appreciate this interaction in other functionalities of the intervention. In the current situation, the user provided information about their recovery progress to the program. However, no interactive feedback was provided on whether the patient was on track, and no follow-up questions were asked in case a patient was not able to keep up. Furthermore, patients felt the need to hear how they were doing. The graphically displayed progress on the website and in the mobile app was not enough; they had a need to hear it from the involved health care professional, which gave them more reassurance. In the current set-up, the doctor was not engaged and up-to-date in the recovery process of the participant at home. Participants valued the opinion of their doctor as very important, and therefore, it is important that the "ikherstel" program is supported by the hospital. They mentioned that this might also prevent conflicting messages in recovery advice:

> *And also some feedback from the hospital, from the treating doctor, the surgeon's assistant that he performs a few calls to check and ask how everything goes and helps a little. I think that, that is the solution is to achieve huge benefits.* [Participant 5, male, 74 years]

Functionality

Participants recognized the added value of the various functionalities. Therefore, it was important that all the functionalities work properly, including the activity tracker and eConsult. Some participants had problems with the activity tracker that was provided with the "ikherstel" program. Even though the activity tracker did not always work properly, participants recognized the added benefit of the concept of an activity tracker and the idea behind it. Therefore, as an alternative for the malfunctioning activity tracker, some participants replaced the provided activity tracker with a built-in step counter on their smartphone or they downloaded an independent activity-tracking app on their mobile phone. This gave them the same opportunity of tracking their activity without the problems that came with the activity tracker that was provided by the "ikherstel" program:

> *The activity tracker was understandable but again that whole synchronization just did not work well. Yes, I did not understand that. I thought that was unfortunate. So I started using my iPhone at a given time. Because that also shows the amount of steps per day.* [Participant 10, male, 64 years]

Participants suggested to include a functionality to check their psychological well-being instead of just focusing on physical recovery. Participants stated that psychological well-being and a positive attitude after cancer surgery influenced recovery and should, therefore, be an integral part of the recovery process. Including a functionality that investigates psychological well-being in the "ikherstel" program would be of added value according to participants.

Participants appreciated the fact that they always had the app at hand and could easily and frequently check their recovery progress during the day. In the beginning, participants used the website to make the personalized recovery plan and read the information on the website, and then, they switched to the app. Participants, therefore, recognized the app as an extra benefit of the "ikherstel" program.

Effort Expectancy

Participants found the "ikherstel" program easy to use, practical, and easy to fit into their daily life with the exception of the activity tracker. Participants stated that it does not cost any effort to use the "ikherstel" program. However, it was mentioned that in the beginning it did cost effort to set up the "ikherstel" program and to find out how it works. Participants suggested that it would be useful if more help would be available for setting up the "ikherstel" program:

> *Practical. Easy. Yes, you could just fit it into your daily life.* [Participant 12, female, 65 years]

Social Influence

There was little or no social influence of family and friends on participants in using the "ikherstel" program. Participants had approached the intervention very individualistically. When participants informed family and friends about the "ikherstel"

program, they were either positive about the fact that the patient participated or they had no opinion about it. Social influence from health care professionals is also quite relevant in this construct. More involvement of health care professionals with the "ikherstel" program is suggested so that patients feel engagement and support from the doctor and hospital:

> *I've seen it has been established and I got the feeling that it was fixed. So it would be nice if someone looks at the recovery monitor and sees that a patient stays behind in certain areas or goes very quickly in certain areas and that the recovery plan can be adjusted accordingly. So that it really becomes interactive.* [Participant 14, male, 48 years]

Facilitating and Inhibiting Conditions

Alternative Course of Disease

A major inhibiting condition for future usage of the "ikherstel" program is the inflexibility of the personalized recovery plan in case of an alternative course of disease. This can happen, for example, in case of complications or additional surgical procedures (ie, removal of a stoma at a later stage). In case the recovery process of the participant took longer, there was currently no option to adjust the personalized recovery plan. Due to this inflexibility, the "ikherstel" program was graded as insufficient by these participants. Participants stated there should be an option to revise the recovery plan in case of an alternative course of disease to their new situation:

> *Some things happened in between, I had an extra hospital admission, as a result I found that "ikherstel" did not fit well with the situation I was in.* [Participant 10, male, 64 years]

Chemotherapy

The "ikherstel" program was also rated insufficient by participants who received chemotherapy in addition to their surgery. For these participants, the surgical procedure was considered more as a side issue, whereas the "ikherstel" program was targeted mainly at recovering from colorectal surgery. Participants who received chemotherapy had a need for additional supportive care after recovery from surgery. They also had a desire for prolonged use of the activity tracker. Participants used the tracker in the first 8 weeks after surgery, and this motivated them enormously. After 6 to 8 weeks, chemotherapy started, but after 8 weeks, the activity tracker had to be sent back to the researcher:

> *Now it was in my opinion much more focused on ehh well you've had an operation. How do you recover from that operation and that kind of thing. But not afterwards of, okay, ehh now there follows chemotherapy, what does that mean and what can we offer in the program to help people. It is certainly helpful.* [Participant 6, male, 64 years]

Discussion

Principal Findings

This process evaluation used both a quantitative and qualitative approach to evaluate the implementation process of a recovery-orientated eHealth intervention. In the quantitative part, average implementation scores per functionality of approximately 60% were reached. A mean score of 7.1 out of 10 for the website, 7.5 for the mobile app, and 7.1 for the activity tracker was given by participants. Barriers and facilitators for use were identified in the qualitative part of this evaluation. The program was experienced by participants as supportive and provided guidance on their recovery progress. Participants had a positive view about the concept and the ease of use of the "ikherstel" program. It was reported that not all functionalities of the program worked properly, and for further sustainable implementation, this needs to be optimized. Interaction with and feedback from health care professionals was a preferred feature by participants, which is currently lacking. In case participants had a different course of recovery, a complication, or received adjuvant chemotherapy, it was impossible to revise or adjust the personalized recovery plan to their new situation. The lack of this function was considered a barrier for use by those with a different course of recovery.

Comparison With Other Studies

There are limited process evaluations concerning recovery-orientated eHealth programs available, making it difficult to compare the results with those of other similar interventions. However, 2 process evaluations performed by our research group regarding a comparable intervention have been published [21,22]. In these studies, similar positive experiences were found. The intervention was found to be supportive, participants had the feeling that they resumed activities quicker, and they found the app to be a convenient and helpful tool. The dose received and fidelity scores for the mobile app and activity tracker confirmed these findings. However, Bouwsma et al reported that some participants experienced the recovery plan being too optimistic for their own situation, whereas others found it too conservative. This is in contrast with the current findings that the recovery plan was achievable or too conservative but not too optimistic [21].

The perceived barrier of limited involvement of health care professionals is in line with findings in other studies [21]. This limited engagement was underlined by the low fidelity score of only 25% of questions asked being answered by health care professionals. A study combined Web-based modules with face-to-face coaching to guide people with early-stage dementia. Participants of this self-management program appreciated the tailored content and positive feedback of the health care professional, which increased the blended structure of the program, resulting in openness of the patients [23]. Another study provided patients a behavior change counseling monitoring and feedback tool with an activity tracker to stimulate physical activity by giving feedback on physical activity performance. This led to more discipline in carrying out activities [24]. These results are in line with our findings that participants indicated that without the activity tracker, they would have been slower in resuming activities, which is an important part of the intervention. However, participants of the program by Boots et al appreciated the personal attention given by the nurses as they combined the activity tracker with counseling in this eHealth program [23]. In both eHealth programs, participants appreciated the personal attention and the blended structure of the

intervention, which was found to be lacking in the eHealth intervention in our study [23,24].

The other major barrier of our intervention was the inability to adjust or adopt the program in case of a different recovery course or additional treatment, resulting in a lack of tailored and personalized care. This is also reflected in the low fidelity score (28%) for the use of the website from those patients. In the study of Bouwsma et al, the intervention was also found to be inflexible in case of complications [21]. It is important for patients recovering after surgery to only receive information that applies to their situation, which is also confirmed in other eHealth-orientated programs [25,26]. When comparing our results with other evaluations regarding interventions for support for cancer survivors, similar results were found. Participants found the intervention to be easily accommodated into their daily routine. However, the need for more specific and tailored information was also considered important in these studies, which is comparable with our findings [27,28].

Strengths and Limitations

A major strength of this study is the mixed-methods approach. Quality in health care is a multidimensional and complex process where some questions and information about quality of care and services are not suitable for quantification. Another strength is that the data collection and analyses were systematically performed using established theoretical frameworks for both the quantitative (the Linnan and Steckler model) and qualitative part (UTAUT framework). Usability of these frameworks was proven in previous process evaluations using these frameworks [15,29]. Another strength is that participants with both low and high intervention ratings were included in the qualitative part, resulting in both positive and critical views of the interventions being equally represented in this evaluation. A limitation of this approach is that patients who stopped prematurely with the intervention were not approached for an interview in the qualitative part because their medical health situation at that time was unknown. Reasons explaining why certain patients reported that the intervention was too difficult could provide information for the construct "effort expectancy." This construct could be used to optimize the convenience for those patients who did not have much experience with information and communication technology or eHealth. Another limitation is the duration between using the intervention and the time the interviews were held. Time varied up to a year, potentially resulting in recall bias where some participants were not able to completely or accurately recall all the details and experiences with the "ikherstel" program. The qualitative data were only collected in a small subsample (n=14) of the study population, and therefore, these data were only presented as an example descriptively and should be interpreted with caution.

Implications for Future Adaptations of the Intervention

To ensure a successful and sustainable implementation of the "ikherstel" program in daily practice, adaptations to the intervention are required. We suggest that health care professionals should be more involved to create more interaction. A more blended care should be introduced to solve the lack in interaction with and feedback from health care professionals. This may increase the engagement of health care professionals in the recovery process after transition to domestic recovery. We note that the hypothesis of this intervention was among others to reduce costs and time spent in the hospital by partially substituting care by this program. We must, therefore, ensure that the use of this eHealth program is additional to traditional care and prevent that care will still be duplicated. We suggest printing a recovery report that patients can take to the outpatient clinic when visiting the surgeon postoperatively to increase involvement of health care professionals. An alternative for receiving feedback would be automatically generated stimulating or confirming push messages to participants with information on their recovery progress.

The personalized recovery plan function has to be adapted so that it is more flexible and can be updated accordingly to the needs of the patient in case of an alternative disease course. The need for less generic and even more tailored information and convalescence advice was considered important by participants. In addition, the "ikherstel" program should be combined with a platform for specialized oncological aftercare for patients who will receive adjuvant chemotherapy. This enables to give patients better advice and support in the period of receiving adjuvant chemotherapy, given their treatment is not finished after the surgical procedure [30].

Clinical Implications

This study underpins the relevance of commitment of health care professionals in supporting eHealth programs. The results reconfirmed the lack of engagement of health care professionals, despite the efforts made to improve user-friendliness and limit the time required for health care professionals. Awareness of health care professionals about eHealth possibilities has to be improved. This is important for implementing this eHealth program, because by supporting this program by health care professionals, patients who will have a colorectal resection will feel more confident and connected with the program. More involvement of health care professionals is also important for future eHealth programs where health care professionals play an important role in public engagement with eHealth and promotion of eHealth functionalities to patients in general. A possible recommendation is to include eHealth as an integral part of residents' education and training. This could be the next step in the implementation of new technologies to the inherently changing health care system.

Conclusions

The "ikherstel"-perioperative care program was experienced as supportive and useful in the recovery process after colorectal surgery. For this indication, the intervention needs more interaction with and feedback from health care professionals and needs more tailored guidance in case of different recovery or treatment courses. To ensure a successful implementation of the "ikherstel" program in daily practice, the awareness and involvement of health care professionals is essential. In our opinion, the recovery process of patients will benefit from an improved blended care approach that bridges the current gap between health care professionals and patients and ensures that eHealth will become part of daily practice.

Acknowledgments

This study is carried out with funding from ZonMw (project number 837002409), an organization for health research and development in the Netherlands.

Authors' Contributions

All authors made substantial contributions to this study. All authors participated in the interpretation of the analyzed data and in the data presentation in this study. In addition, all authors participated in the drafting and revising of this study, and all the authors approve this version to be published.

Conflicts of Interest

CMdB, FGS, and CdG have no conflicts of interest. JRA and JAFH intend to set up a spin-off company concerning the implementation of a mobile app concerning the IKHERSTEL intervention in the Netherlands. JAFH received grants from NWO, ZonMw, and Samsung during the conduct of the study and received a fee from Olympus, outside the submitted work. HJB received personal fees from Olympus, Stryker, Medtronic, and Applied medical, outside the submitted work. JRA holds a chair in Insurance Medicine paid by the Dutch Social Security Agency; he is a stockholder of Evalua and received grants from ZonMw/NWO, InstituutGak, UWV, SZW, VWS, Pfizer, Achmea, and CVZ/ZorgInstituut, outside the submitted work.

Multimedia Appendix 1

Patient characteristics of the qualitative part of the study.

[XLSX File (Microsoft Excel File), 11KB - jmir_v21i1e10674_app1.xlsx]

Multimedia Appendix 2

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 434KB - jmir_v21i1e10674_app2.pdf]

References

1. Spanjersberg WR, Reurings J, Keus F, van Laarhoven CJ. Fast track surgery versus conventional recovery strategies for colorectal surgery. Cochrane Database Syst Rev 2011 Feb 16(2):CD007635. [doi: 10.1002/14651858.CD007635.pub2] [Medline: 21328298]
2. Bonjer HJ, Deijen CL, Abis GA, Cuesta MA, van der Pas MH, de Lange-de Klerk ES, COLOR II Study Group. A randomized trial of laparoscopic versus open surgery for rectal cancer. N Engl J Med 2015 Apr 02;372(14):1324-1332. [doi: 10.1056/NEJMoa1414882] [Medline: 25830422]
3. Dutch Hospital data. 2013 Dec. Kengetallen Nederlandse Ziekenhuizen 2012 URL: https://www.nvz-ziekenhuizen.nl/_library/13152/Kengetallen%20Nederlandse%20Ziekenhuizen%202012.pdf [accessed 2018-11-30] [WebCite Cache ID 74J9G1Uxn]
4. Henderson M, Glozier N, Holland EK. Long term sickness absence. Br Med J 2005 Apr 09;330(7495):802-803 [FREE Full text] [doi: 10.1136/bmj.330.7495.802] [Medline: 15817531]
5. Lewis J, Ray P, Liaw S. Recent worldwide developments in eHealth and mHealth to more effectively manage cancer and other chronic diseases - a systematic review. Yearb Med Inform 2016 Nov 10(1):93-108 [FREE Full text] [doi: 10.15265/IY-2016-020] [Medline: 27830236]
6. Barello 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]
7. Tang C, Lorenzi N, Harle CA, Zhou X, Chen Y. Interactive systems for patient-centered care to enhance patient engagement. J Am Med Inform Assoc 2016 Jan 14;23(1):2-4. [doi: 10.1093/jamia/ocv198]
8. Dedding C, van Doorn R, Winkler L, Reis R. How will e-health affect patient participation in the clinic? A review of e-health studies and the current evidence for changes in the relationship between medical professionals and patients. Soc Sci Med 2011 Jan;72(1):49-53. [doi: 10.1016/j.socscimed.2010.10.017] [Medline: 21129832]
9. van der Meij E, Huirne JA, Bouwsma EV, van Dongen JM, Terwee CB, van der Ven PM, et al. Substitution of usual perioperative care by eHealth to enhance postoperative recovery in patients undergoing general surgical or gynecological procedures: study protocol of a randomized controlled trial. JMIR Res Protoc 2016 Dec 21;5(4):e245 [FREE Full text] [doi: 10.2196/resprot.6580] [Medline: 28003177]
10. Hasson H. Systematic evaluation of implementation fidelity of complex interventions in health and social care. Implement Sci 2010 Sep 03;5:67 [FREE Full text] [doi: 10.1186/1748-5908-5-67] [Medline: 20815872]

11. Munro A, Bloor M. Process evaluation: the new miracle ingredient in public health research? Qual Res 2010;10:699-713. [doi: 10.1177/1468794110380522]
12. McGraw SA, Sellers DE, Stone EJ, Bebchuk J, Edmundson EW, Johnson CC, et al. Using process data to explain outcomes: an illustration from the child and adolescent trial for cardiovascular health (CATCH). Eval Rev 1996 Jun;20(3):291-312. [doi: 10.1177/0193841X9602000304] [Medline: 10182206]
13. Abelson JS, Kaufman E, Symer M, Peters A, Charlson M, Yeo H. Barriers and benefits to using mobile health technology after operation: a qualitative study. Surgery 2017 Dec;162(3):605-611. [doi: 10.1016/j.surg.2017.05.007] [Medline: 28651777]
14. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. Br Med J 2015;350:h1258 [FREE Full text] [Medline: 25791983]
15. Suman A, Schaafsma FG, Bamarni J, van Tulder MW, Anema JR. A multimedia campaign to improve back beliefs in patients with non-specific low back pain: a process evaluation. BMC Musculoskelet Disord 2017 Dec 18;18(1):200 [FREE Full text] [doi: 10.1186/s12891-017-1551-z] [Medline: 28521761]
16. 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] [Medline: 22209829]
17. den Bakker CM, Schaafsma FG, van der Meij E, Meijerink W, van den Heuvel B, Baan A. eHealth program to empower patients in returning to normal activities after general surgical and gynecological procedures: Intervention Mapping as a useful method for further development. J Med Internet Res 2018:- Epub ahead of print(forthcoming). [doi: 10.2196/jmir.9938]
18. Linnan L, Steckler A. Process Evaluation for Public Health Interventions and Research. 1st edition. San Francisco: Jossey-Bass; 2001.
19. Venkatesh V, Thong J, Xu X. Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. JAIS 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
20. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822X05279903] [Medline: 22973420]
21. Bouwsma EV, Vonk Noordegraaf A, Szlávik Z, Brölmann HA, Emanuel MH, Lips JP, et al. Process evaluation of a multidisciplinary care program for patients undergoing gynaecological surgery. J Occup Rehabil 2014 Sep;24(3):425-438 [FREE Full text] [doi: 10.1007/s10926-013-9475-4] [Medline: 24057871]
22. van der Meij E, Huirne JA, Ten Cate AD, Stockmann HB, Scholten PC, Davids PH, et al. A perioperative eHealth program to enhance postoperative recovery after abdominal surgery: process evaluation of a randomized controlled trial. J Med Internet Res 2018 Jan 02;20(1):e1 [FREE Full text] [doi: 10.2196/jmir.8338] [Medline: 29295808]
23. 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] [Medline: 29258980]
24. Verwey R, van der Weegen S, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. Process evaluation of physical activity counselling with and without the use of mobile technology: a mixed methods study. Int J Nurs Stud 2016 Jan;53:3-16. [doi: 10.1016/j.ijnurstu.2015.10.008] [Medline: 26518108]
25. Martínez-Ramos C, Cerdán MT, López RS. Mobile phone-based telemedicine system for the home follow-up of patients undergoing ambulatory surgery. Telemed J E Health 2009;15(6):531-537. [doi: 10.1089/tmj.2009.0003] [Medline: 19566396]
26. Semple JL, Sharpe S, Murnaghan ML, Theodoropoulos J, Metcalfe KA. Using a mobile app for monitoring post-operative quality of recovery of patients at home: a feasibility study. JMIR Mhealth Uhealth 2015;3(1):e18 [FREE Full text] [doi: 10.2196/mhealth.3929] [Medline: 25679749]
27. Myall M, May CR, Grimmett C, May CM, Calman L, Richardson A, et al. RESTORE: an exploratory trial of a web-based intervention to enhance self-management of cancer-related fatigue: findings from a qualitative process evaluation. BMC Med Inform Decis Mak 2015 Nov 14;15:94 [FREE Full text] [doi: 10.1186/s12911-015-0214-y] [Medline: 26577690]
28. Kanera IM, Willems RA, Bolman CA, Mesters I, Zambon V, Gijsen BC, et al. Use and appreciation of a tailored self-management eHealth intervention for early cancer survivors: process evaluation of a randomized controlled trial. J Med Internet Res 2016;18(8):e229 [FREE Full text] [doi: 10.2196/jmir.5975] [Medline: 27554525]
29. Hennemann S, Beutel ME, Zwerenz R. Drivers and barriers to acceptance of web-based aftercare of patients in inpatient routine care: a cross-sectional survey. J Med Internet Res 2016 Dec 23;18(12):e337 [FREE Full text] [doi: 10.2196/jmir.6003] [Medline: 28011445]
30. Duman-Lubberding S, van Uden-Kraan CF, Jansen F, Witte BI, van der Velden LA, Lacko M, et al. Feasibility of an eHealth application "OncoKompas" to improve personalized survivorship cancer care. Support Care Cancer 2016 May;24(5):2163-2171 [FREE Full text] [doi: 10.1007/s00520-015-3004-2] [Medline: 26563178]

Abbreviations

eConsult: electronic consult
eHealth: electronic health

RCT: randomized controlled trial
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by G Eysenbach; submitted 03.04.18; peer-reviewed by G Kok, W Leclercq; comments to author 04.07.18; revised version received 12.07.18; accepted 16.07.18; published 29.01.19.

Please cite as:
den Bakker CM, Huirne JAF, Schaafsma FG, de Geus C, Bonjer HJ, Anema JR
Electronic Health Program to Empower Patients in Returning to Normal Activities After Colorectal Surgical Procedures: Mixed-Methods Process Evaluation Alongside a Randomized Controlled Trial
J Med Internet Res 2019;21(1):e10674
URL: http://www.jmir.org/2019/1/e10674/
doi:10.2196/10674
PMID:30694205

Original Paper

Using the Facebook Advertisement Platform to Recruit Chinese, Korean, and Latinx Cancer Survivors for Psychosocial Research: Web-Based Survey Study

William Tsai[1], PhD; Daisy Zavala[2], BA; Sol Gomez[2], BA

[1]Department of Applied Psychology, New York University, New York, NY, United States

[2]Department of Psychology, California State University San Marcos, San Marcos, CA, United States

Corresponding Author:
William Tsai, PhD
Department of Applied Psychology
New York University
246 Greene Street, 8th Floor
New York, NY, 10003
United States
Phone: 1 212 998 5552
Email: will.tsai@nyu.edu

Abstract

Background: Ethnic minority cancer survivors remain an understudied and underrepresented population in cancer research, in part, due to the challenge of low participant recruitment rates. Therefore, identifying effective recruitment strategies is imperative for reducing cancer health disparities among this population. With the widespread use of social media, health researchers have turned to Facebook as a potential source of recruitment.

Objective: We aimed to evaluate the feasibility and effectiveness of purchasing ads on Facebook to recruit Chinese, Korean, and Latinx cancer survivors residing in the United States. We assessed their experience with participating in a Web-based survey and their interest for future research.

Methods: We showed 5 purchased ads in English, simplified Chinese, traditional Chinese, Korean, and Spanish on Facebook. Participants who clicked on the Facebook ad were directed to the study website and asked to submit their emails to receive the link to the 30-minute Web-based survey. Inclusion criteria included being of Asian or Latinx heritage, age ≥18 years, having a cancer diagnosis, and being within 5 years of cancer treatment. Participants who completed the survey were sent a US $10 Walmart eGiftcard.

Results: The Facebook ads were shown for 48 consecutive days for a total spending of US $1200.46 (US $25/day budget). Overall, 11 East Asian and 15 Latinx cancer survivors completed the study, resulting in an average cost per participant of US $46.17. The East Asian and Latinx cancer survivors did not significantly differ in age, years lived in the United States, education level, generation status, and time since diagnosis. However, Latinx cancer survivors were marginally more likely to have limited English proficiency and lower annual income than East Asian cancer survivors. Both Latinx and East Asian cancer survivors reported that they enjoyed participating in this study and indicated an interest in participating in future psychosocial research studies.

Conclusions: The use of Facebook ads successfully resulted in the recruitment of East Asian and Latinx cancer survivors with different cancer diagnoses who reside in various geographic regions of the United States. We found that East Asian and Latinx cancer survivors recruited through Facebook were interested in participating in future psychosocial research, thereby providing support for the feasibility and effectiveness of using Facebook as a source of recruitment for ethnic minority cancer survivors.

(J Med Internet Res 2019;21(1):e11571) doi:10.2196/11571

KEYWORDS

ethnic minority cancer survivor; Facebook recruitment; Korean cancer survivor; Chinese cancer survivor; Latinx cancer survivor; mobile phone

Introduction

Because ethnic minority cancer survivors experience cancer health disparities along the entire cancer control continuum [1], initiatives such as the Healthy People 2020 [2] have called for greater attention and efforts to increase the representation of ethnic minorities in cancer research. However, they remain an understudied and underrepresented population in cancer research in part owing to the persistent challenge of low recruitment rates. Despite researchers' persistent efforts, traditional recruitment strategies such as utilizing physician referrals, purchasing newspaper ads, and using cancer registries have resulted in mixed results [3,4]. The low recruitment rates of ethnic minority cancer survivors in cancer research remain a concern, and the low rates of participation also call into question the generalizability and applicability of extant findings and interventions for this population. There are substantial public health and societal benefits to be gained by identifying culturally sensitive and effective recruitment strategies for this population.

Asian American people are the fastest growing and Latinx American people are the largest ethnic minority population in the United States. Despite marked improvements in cancer prognosis and survival rates for the general population, Asian American and Latinx American people are more likely to experience poorer quality of life relative to non-Hispanic white people [5]. Recent research has also showed increases in the annual incidence rate of breast cancer among Asian American women [6]. There are many reasons for the low participation rate of ethnic minority cancer survivors in research. These reasons may include language barriers [7], low socioeconomic status, lack of health insurance and access to care [8], cultural barriers such as distrust of researchers [9], and stigmatized beliefs about cancer [10]. These barriers limit access to traditional recruitment methods that often rely on physicians to make referrals in health care settings.

With the widespread use of social media, health researchers have turned to Facebook as a potential source of recruitment [11]. As the most popular social media website, with 2.2 billion active users who have logged in at least once during the last 30 days [12], Facebook has qualities that may help overcome barriers and facilitate the recruitment of ethnic minority cancer survivors. For instance, the language barrier can be overcome by translating ads into different languages and targeting them to appear to users who are using Facebook in their native languages. With a potential to overcome barriers to recruitment in health care settings, there is evidence that Facebook is now widely used by adults aged ≥65 years and those with low socioeconomic status backgrounds [13]. Specifically, 62% of older adults who used the internet used Facebook, and 77% of people with an annual household income <US $30,000 used Facebook in 2016 [14,15]. These trends suggest that Facebook has a strong potential to reach ethnic minority cancer survivors who come from lower socioeconomic status backgrounds that may not otherwise have exposure to research opportunities. Lastly, in the privacy of their homes, potential participants who are less trusting of researchers can take the time they need to evaluate the benefits and costs of participating in research.

An increasing number of studies have demonstrated the feasibility of using Facebook to recruit participants for health and medical research [11]. However, to our knowledge, we are the first to evaluate the feasibility and effectiveness of paid advertisements on Facebook as a platform for identifying Chinese, Korean, and Latinx American cancer survivors. We assessed participants' experience with participating in this Web-based survey study and their willingness to participate in future research studies.

Methods

Facebook Ads

In total, we purchased 5 ads in English, simplified Chinese, traditional Chinese, Korean, and Spanish and showed them on Facebook from June 16 to August 3, 2017. The California State University San Marcos Institutional Review Board approved the ads shown on Facebook. Following the recommendations by Thornton et al [11], the ads were concise, showed the affiliation of the research study with a university, and offered incentives for participation (Figure 1). The link in each Facebook ad connected to the study website, which stated the eligibility criteria for participating in the 30-minute Web-based survey in the respective language of the clicked ads (eg, clicking the Korean ad would lead to the study information presented in Korean).

Study Procedures

When participants clicked on the Facebook ad, they were directed to the study website that provided a general description of the study and the inclusion criteria. On this website, participants interested in the study submitted their email address by entering it into the "email" submission box. Then, participants received a more detailed description of the study as well as a link to the Web-based survey via email. We directed participants who accessed the survey link from their email inbox to the Web-based survey cover page that described the study, which explained the inclusion criteria, study procedure, and data confidentiality and stated that their consent was given by continuing with the survey. The first 5 questions screened potential participants for eligibility by asking whether they were Asian or Latinx, lived in the United States, aged ≥18 years, diagnosed with cancer, and within 5 years of their cancer treatment. Participants who answered "yes" to each of the 5 questions were directed to complete the survey, which assessed their demographic and cancer diagnosis information, well-being, experience with the study, and interest in participating in future research studies. We did not report the data on well-being in this study. Participants who completed the survey were sent a US $10 Walmart eGiftcard.

Figure 1. Example of Facebook ads. The simplified Chinese ad is not shown.

Measures

Participation Experience

We assessed participants' experience with the Web-based research study using 3 items rated on a 4-point Likert Scale ranging from 1 ("Strongly disagree") to 4 ("Strongly agree"). We assessed the perceived difficulty with completing the research study using 2 items (ie, "I did not experience technical difficulty while participating in this research" and "Participating in this study was easy"). A composite perceived difficulty variable was created by averaging these 2 items. The third item assessed the extent to which they enjoyed having the opportunity to share their feelings and experiences with cancer through this research study (ie, "I enjoyed the opportunity to share my feelings and experience about cancer in this research study"). Higher scores indicate lower perceived difficulty with completing the research and greater enjoyment of having completed the survey.

Willingness and Access to Smartphone to Participate in Future Research

We assessed participants' interest in future research studies using 4 items (eg, "I would like to participate in more research studies in the future") rated on a 4-point Likert Scale ranging from 1 ("Strongly disagree") to 4 ("Strongly agree"). The other 3 items assessed participants' willingness to participate in a Web or phone interview with a research staff member, willingness to privately write about their cancer experiences, and willingness to privately voice record their cancer experiences. The 4 items were averaged, with higher scores indicating greater willingness to participate in future research (Cronbach alpha=.80 and.88 for Latinx and Asian cancer survivors, respectively). Lastly, we assessed whether the

participant had a smartphone, as a means to participate in future research.

Advertising Campaign

The Facebook advertisement platform (ie, Facebook Ads Manager) provides researchers with the ability to target the age, location, interests, and behaviors of those who will be shown the ads. We targeted ads at Facebook users with interests in cancer-related keywords or organizations (eg, American Cancer Society, Oncology, Breast cancer care, Skin cancer or Melanoma awareness, National Cancer Institute, Susan G. Komen for the Cure, cancer awareness, and American Cancer Society Relay for Life) and behaviors related to cancer causes. In addition to the target interests and behaviors above, the English ad targeted participants who used Facebook in English, simplified or traditional Chinese, and Spanish and those who indicated interests in Hispanic or Asian culture. For the other translated ads, they were delivered to Facebook users who used that specific language and those who indicated interests or behaviors with an affinity to the respective culture (eg, the Spanish ad was delivered to participants who used Facebook in Spanish and those who indicated interests or behaviors related to Hispanic culture). All ads were delivered to adult Facebook users who lived in the United States.

Results

Advertisement Campaign

The ad campaign was set up with a maximum daily budget of US $25 and ran for 48 consecutive days for a total spending of US $1,200.46 (June 16 to August 3, 2017). In total, the English, Spanish, Korean, and traditional and simplified Chinese ads generated 45,768, 34,261, 30,335, 21,230, and 13,041 impressions (ie, total number of times the ads were shown to Facebook users) and reached 31,299, 25,217, 8187, 7183, and 4915 people (ie, number of unique Facebook users who saw the ads), respectively. These impressions generated 420, 984, 307, 291, and 127 link clicks for the English, Spanish, Korean, and traditional and simplified Chinese ads, respectively. The average cost per click across the 5 ads was US $0.56 (see Table 1). Of the 2129 total link clicks that directed the potential participant to the study website, 115 people entered their emails after reading the study information and eligibility criteria (Figure 2). Due to 10 incorrectly submitted emails, 10 survey links were not delivered. Of the105 people who were sent the Web-based survey link, 36 (34.3%) clicked on the survey link and completed the initial screening questions. Of these 36 people who begun the survey, 29 (81%) were eligible and completed the Web-based survey (10 Chinese, 4 Korean, and 15 Latinx cancer survivors); however, 3 (10%) surveys were completed by a Chinese participant with the same internet protocol address and, thus, were excluded from subsequent analyses. In total, data from 4 surveys in English (3 by Latinx and 1 by Chinese participants), 4 surveys in Korean, 5 surveys in traditional Chinese, 1 survey in simplified Chinese, and 12 surveys in Spanish were analyzed. The cost per participant was US $46.17 (ie, total spending per number of completed Web-based surveys).

Participant Demographic and Cancer Diagnosis Information

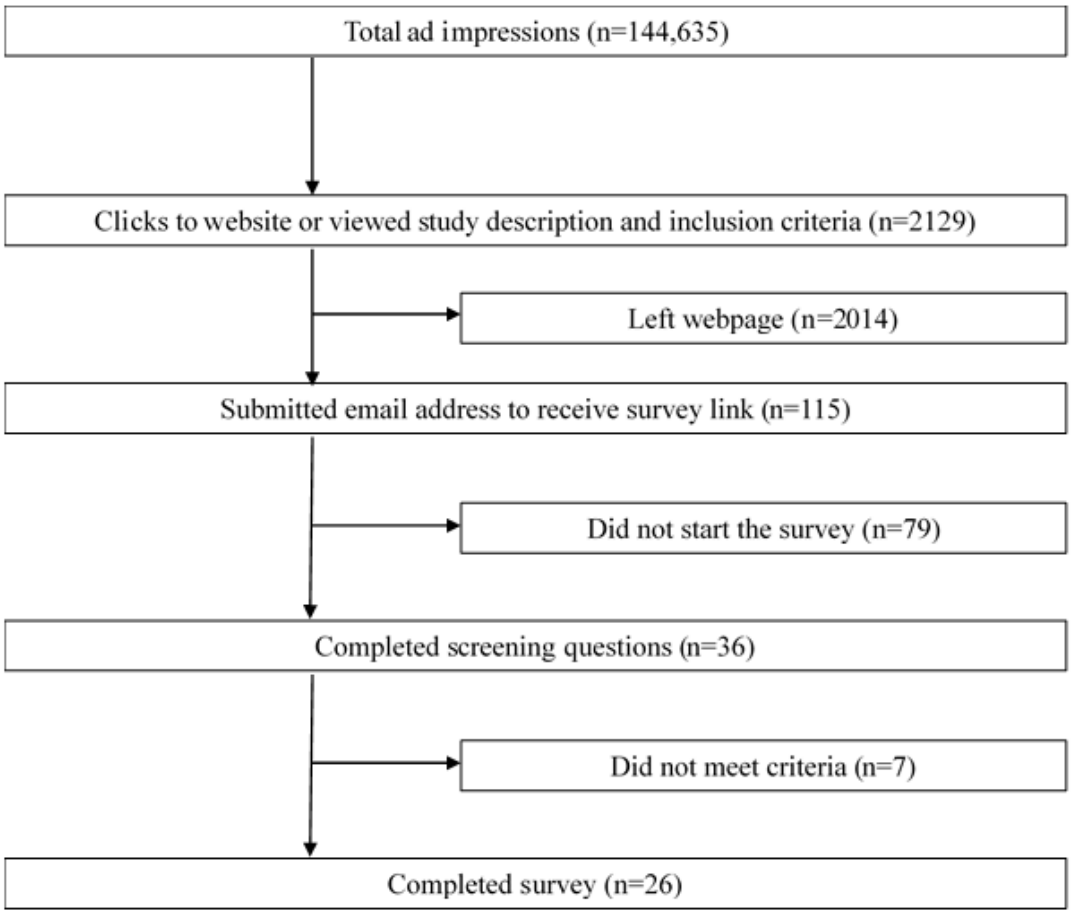
We combined the Chinese and Korean participants into 1 East Asian participant group. Of the 11 East Asian cancer survivors (mean age 55.91 [SD 10.03] years; 8/11, 73%, females), 91% (10/11) were first-generation immigrants (ie, foreign-born), 55% (6/11) obtained a college degree, and 9% (1/11) reported an annual family income <US $13,000. The majority (9/11, 82%) spoke Chinese or Korean but 64% (7/11) also reported the ability to speak English. Most of the East Asian cancer survivors had diagnoses of breast cancer (5/11, 46%) or lymphoma (3/11, 27%), and 55% (6/11) were in remission. Of the 15 Latinx cancer survivors (mean age 49.73 [SD 9.71] years; 13/15, 87%, females), 80% (12/15) were first-generation immigrants, 53% (8/15) obtained a college degree, and 47% (7/15) reported an annual family income <US $13,000. The majority (13/15, 87%) spoke Spanish, but only 20% (3/15) also reported the ability to speak English. Most of the Latinx cancer survivors had diagnoses of breast cancer (7/15, 47%) or colorectal cancer (3/15, 20%). See Table 2 for complete demographic and cancer diagnosis information.

Table 1. Facebook advertisement campaign.

Ad language	Link clicks	Impressions (people)	Reach (people)	Cost per click (US $)	Amount spent (US $)
English	420	45,768	31,299	0.67	281.37
Spanish	984	34,261	25,217	0.31	307.97
Korean	307	30,335	8187	0.97	296.29
Simplified Chinese	127	13,041	4915	0.66	83.60
Traditional Chinese	291	21,230	7183	0.79	231.23
Total	2129	144,635	71,412[a]	0.56	1200.46

[a]This number reflects the total unique number of people who saw at least one ad.

Figure 2. Flow diagram of impressions, ad clicks, email submissions, survey participation, and completion; 3 surveys were completed by a participant with the same internet protocol address and subsequently removed from analyses.



We examined demographic differences between East Asian and Latinx cancer survivors. They did not significantly differ regarding age (*P*=.13), years lived in the United States (*P*=.10), education level (*P*=.92), generation status (*P*>.99), stage of cancer diagnosis (*P*=.35), time since diagnosis (*P*=.85), and treatment status (*P*=.82). However, Latinx cancer survivors were marginally more likely to only speak their heritage language (ie, limited English proficiency; *P*=.06) and marginally more likely to endorse an annual family income <US $13,000 compared with the East Asian cancer survivors (*P*=.09).

Experience with the Web-Based Survey and Interest in Future Research

We examined participants' experience with this study and their interest in participating in future research (Table 3). East Asian cancer survivors reported a lower ease of participation (eg, having technical difficulties accessing the survey) in this study than Latinx cancer survivors (mean 2.91 [SD 0.66] vs mean 3.47 [SD 0.52]; two-tailed t_{24}=2.41; *P*=.02). Both Latinx and East Asian cancer survivors reported that they enjoyed the opportunity to share their cancer-related experiences in this Web-based survey (mean 3.40 [SD 0.63] and mean 3.27 [SD 0.47], respectively; two-tailed t_{24}=0.37; *P*=.56). Both Latinx and East Asian cancer survivors indicated a willingness to participate in future cancer research studies through video or phone interviews with study researchers or sharing their cancer experiences through writing or voice recording (mean 3.00 [SD 0.54] and mean 3.00 [SD 0.63], respectively; two-tailed t_{24}=0; *P*>.99). The majority of Latinx and East Asian cancer survivors indicated having access to a smartphone (13/15, 87%, and 10/11, 91%, respectively).

Table 2. Latinx and Asian cancer survivor demographic and cancer diagnosis information.

Variable	Latinx survivors (n=15)	East Asian survivors (n=11)	*P* value[a]
Ethnicity, n (%)			
Mexican	8 (53)	—[b]	—
Cuban	2 (13)	—	—
Puerto Rican	4 (27)	—	—
Honduran	1 (7)	—	—
Mainland Chinese	—	3 (27)	—
Taiwan Chinese	—	3 (27)	—
Hong Kong Chinese	—	1 (9)	—
Korean	—	4 (36)	—
Age (years), mean (SD)	49.73 (9.71)	55.91 (10.03)	.13
Time since diagnosis (months), mean (SD)	26.36 (12.64)	27.80 (25.10)	.85
Gender, n (%)			.62
Male	2 (13)	3 (27)	—
Female	13 (87)	8 (73)	—
Education, n (%)			.92
Less than high school	3 (20)	1 (9)	—
High school degree	2 (13)	1 (9)	—
Some college	1 (7)	2 (18)	—
College degree	8 (53)	6 (55)	—
Professional or graduate degree	1 (7)	1 (9)	—
Geographic location, n (%)			.25
South	7 (47)	2 (18)	—
East	2 (13)	1 (9)	—
Midwest	0 (0)	2 (18)	—
Northeast	2 (13)	2 (18)	—
West	4 (27)	4 (36)	—
Generation status, n (%)			
First generation	12 (80)	10 (91)	>.99
Second generation	2 (13)	1 (9)	—
Fourth generation	1 (7)	0 (0)	—
Language spoken, n (%)			.06
Heritage language only	12 (80)	4 (36)	—
English only	2 (13)	2 (18)	—
Both heritage and English	1 (7)	5 (45)	—
Marital status, n (%)			.60
Single	2 (13)	1 (9)	—
Married	10 (67)	7 (64)	—
Divorced	1 (7)	2 (18)	—
Separated	2 (13)	0 (0)	—
Bereaved	0 (0)	1 (9)	—
Annual household income, n (%)			.09

Variable	Latinx survivors (n=15)	East Asian survivors (n=11)	*P* value[a]
Less than US $13k	7 (47)	1 (9)	—
US $13k to $30k	2 (13)	2 (18)	—
US $30k to $60k	4 (27)	3 (27)	—
US $60k to $100k	0 (0)	3 (27)	—
Greater than US $100k	1 (7)	2 (18)	—
Missing	1 (7)	0 (0)	—
Stage of cancer diagnosis, n (%)			.35
I	4 (27)	1 (9)	—
II	6 (40)	3 (27)	—
III	2 (13)	2 (18)	—
IV	0 (0)	2 (18)	—
Missing	3 (20)	3 (27)	—
Cancer type, n (%)			.23
Breast	7 (47)	5 (45)	—
Blood	1 (7)	1 (9)	—
Colorectal	3 (20)	0 (0)	—
Oral	1 (7)	0 (0)	—
Skin	1 (7)	0 (0)	—
Thyroid	2 (13)	1 (9)	—
Lymphoma	0 (0)	3 (27)	—
Pancreatic	0 (0)	1 (9)	—
Treatment status, n (%)			.82
Currently in treatment	5 (33)	5 (45)	—
In remission	9 (60)	6 (55)	—
Missing	1 (7)	0 (0)	—
Cancer treatment, n (%)			.91
Chemotherapy only	2 (13)	1 (9)	—
Radiation only	2 (13)	0 (0)	—
Surgery only	2 (13)	0 (0)	—
Medication only	1 (7)	2 (18)	—
Chemotherapy, radiation, and surgery	5 (33)	5 (45)	—
Chemotherapy and radiation	1 (7)	1 (9)	—
Chemotherapy and surgery	1 (7)	1 (9)	—
Radiation and surgery	1 (7)	1 (9)	—

[a]*P* values for categorical variables are based on chi-square or Fisher's exact test, and *P* values for continuous variables are based on two-tailed *t* tests.
[b]Not applicable.

Table 3. Participant experience with this study and willingness to participate in future research.

Participant experience	Latinx (n=15)	East Asian (n=11)	*P* value[a]
Perceived difficulty with study composite[b], mean (SD)	3.47 (0.52)	2.91 (0.66)	.02
I did not experience technical difficulty while participating in this research.	3.40 (0.63)	2.82 (0.87)	.06
Participating in this study was easy.	3.53 (0.52)	3 (0.78)	.05
Participation enjoyment[b], mean (SD)			
I enjoyed the opportunity to share my feelings and experience about cancer in this research study.	3.40 (0.63)	3.27 (0.47)	.56
Interest in future research composite[b], mean (SD)	3.07 (0.43)	3.00 (0.50)	.72
I would like to participate in more research studies in the future.	3.47 (0.52)	3.27 (0.47)	.34
I would be willing to participate in a Web or phone interview with research staff.	2.93 (0.46)	3.09 (0.54)	.43
I would be willing to write about my cancer experiences in a private journal.	3 (0.54)	3 (0.63)	>.99
I would be willing to speak about my cancer experiences by recording into a voice recorder.	2.87 (0.64)	2.64 (0.67)	.38
Access to a smartphone, n (%)			.74
Yes	13 (87)	10 (91)	—[c]
No	2 (13)	1 (9)	—

[a]*P* values for categorical variables are based on chi-square or Fisher's exact test, and *P* values for continuous variables are based on two-tailed *t* tests.

[b]Likert ratings range from 1="Strongly disagree" to 4="Strongly agree".

[c]Not applicable.

Discussion

Principal Findings

This study demonstrated the feasibility and effectiveness of recruiting Latinx and East Asian cancer survivors via targeted advertisements on Facebook. With a focus on identifying cancer survivors who may benefit from participating in psychosocial intervention research (ie, within 5 years of cancer diagnosis), we examined their experience with our Web-based survey and their interest in participating in future research.

We found that the Facebook campaign was successful in recruiting a sample of Latinx and East Asian cancer survivors with different cancer diagnoses from various geographic regions of the United States. Latinx and East Asian cancer survivors recruited from Facebook enjoyed sharing their cancer experiences and reported their interest in participating in future psychosocial intervention studies ranging from Web-based focus groups to private emotional disclosure through writing or voice recording (eg, expressive writing interventions [16]). Furthermore, we found that all but 3 (23/26, 88%) participants had access to smartphones and, thus, could participate in Web-based or mobile-based psychosocial interventions that have become increasingly popular in recent years [17].

Most of the Latinx and East Asian cancer survivors were immigrants, female, and had breast cancer diagnoses. However, Latinx cancer survivors were more likely to speak only their heritage language (eg, Spanish) and reported a lower annual household income than the East Asian cancer survivors. These findings suggest that Facebook may be an especially effective source of recruitment for immigrant Latinx cancer survivors who are from lower socioeconomic backgrounds. Because this population represents a significantly understudied and underserved population with significant cancer health disparities [5], our findings build upon a growing number of studies that have demonstrated the success of using Facebook to recruit harder-to-reach medical populations [18].

A significant portion of potential participants did not participate in the Web-based survey after receiving the survey link in their emails. Although speculative, these participants may have submitted their emails on the spur of the moment after clicking the Facebook advertisement without carefully reading the study description and inclusion criteria on the website, and only later learned from the received email that they were ineligible. Alternatively, it is also possible that they had difficulty trusting the researchers and the email they received [9]. Although we included the affiliated university logo where possible (eg, on the Facebook ad and study website), the research team's email used to send the survey link was not from a university-affiliated email address. Perhaps using a university-affiliated email address would have increased the participants' trust and willingness to complete the survey. Nonetheless, we successfully recruited 11 East Asian and 15 Latinx cancer survivors in 48 consecutive days of recruitment. Notably, we only recruited through the paid Facebook advertisement platform and did not utilize other opportunities that were freely available on Facebook. For example, we did not manually post our ads on cancer-related Facebook pages and groups [19]. Researchers considering Facebook to recruit ethnic minority cancer survivors can potentially maximize their success by utilizing both free and paid advertisement strategies.

Limitations

While we found support for the feasibility and effectiveness of recruiting East Asian and Latinx cancer survivors from Facebook, our study has several limitations. First, we did not include a direct comparison with traditional methods of recruitment (eg, physician referrals or flyering), so we could not determine whether Facebook recruitment is a more efficient or cost-effective method than traditional methods. However, the costs of hiring research staff with language abilities to distribute flyers and screen potential participants are likely more expensive and time-intensive than the minimal time and personnel needed for recruitment through Facebook. Although we cannot speculate on the demographic similarities or differences between the East Asian and Latinx cancer survivors recruited through Facebook or in person, a growing number of studies have found minimal differences between Web-based and in-person recruitments [11]. However, the feasibility and effectiveness of recruiting through other social media platforms (eg, Craigslist) requires future testing. Moreover, given that the study was conducted through the Web, we could not verify the veracity of the information provided by the participants. It is also unclear whether participants who clicked on the ads and viewed the study page were eligible but not interested in participating or whether they were ineligible owing to having completed cancer treatment over 5 years ago. Lastly, given the focus on Chinese and Korean cancer survivors in this study, the generalizability of our findings to other Asian subgroups (eg, Filipino and Hmong) from Facebook remains to be tested.

Conclusions

The recruitment of ethnic minority cancer survivors into cancer research has been a longstanding issue. With the potential to address this problem, we found that Facebook was a feasible and effective platform for recruiting Latinx and East Asian cancer survivors who may be interested in future psychosocial intervention research. Facebook was successful in reaching participants with different cancer diagnoses who come from various geographic regions. Ethnic minority cancer survivors are regarded as a challenging population to recruit, partially owing to limited English proficiency and low socioeconomic status. The Latinx and East Asian cancer survivors who participated in this study reported favorable experiences with the Web-based survey and reported interest in participating in future research studies. Future research studies should compare demographic and psychosocial differences between participants recruited from traditional and Web-based sources.

Acknowledgments

This work was supported by the California State University San Marcos Grant Proposal Seed Money Award.

Conflicts of Interest

None declared.

References

1. Kagawa-Singer M, Dadia AV, Yu MC, Surbone A. Cancer, culture, and health disparities: time to chart a new course? CA Cancer J Clin 2010;60(1):12-39 [FREE Full text] [doi: 10.3322/caac.20051] [Medline: 20097836]
2. Healthy People 2020. URL: http://www.healthypeople.gov/ [WebCite Cache ID 6yiMHReux]
3. Joseph G, Dohan D. Recruiting minorities where they receive care: Institutional barriers to cancer clinical trials recruitment in a safety-net hospital. Contemp Clin Trials 2009 Nov;30(6):552-559. [doi: 10.1016/j.cct.2009.06.009] [Medline: 19580887]
4. Ganz PA, Land SR, Antonio C, Zheng P, Yothers G, Petersen L, et al. Cancer survivorship research: the challenge of recruiting adult long term cancer survivors from a cooperative clinical trials group. J Cancer Surviv 2009 Sep;3(3):137-147 [FREE Full text] [doi: 10.1007/s11764-009-0093-2] [Medline: 19526347]
5. American Cancer Society. 2014. Cancer Facts & Figures 2014 URL: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2014.html [WebCite Cache ID 6yiPWGOe6]
6. Gomez SL, Von Behren J, McKinley M, Clarke CA, Shariff-Marco S, Cheng I, et al. Breast cancer in Asian Americans in California, 1988-2013: increasing incidence trends and recent data on breast cancer subtypes. Breast Cancer Res Treat 2017 Jul;164(1):139-147 [FREE Full text] [doi: 10.1007/s10549-017-4229-1] [Medline: 28365834]
7. Yi JK, Swartz MD, Reyes-Gibby CC. English proficiency, symptoms, and quality of life in Vietnamese- and Chinese-American breast cancer survivors. J Pain Symptom Manage 2011 Jul;42(1):83-92. [doi: 10.1016/j.jpainsymman.2010.09.014] [Medline: 21227634]
8. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annu Rev Public Health 2006;27:1-28. [doi: 10.1146/annurev.publhealth.27.021405.102113] [Medline: 16533107]
9. Moreno-John G, Gachie A, Fleming CM, Nápoles-Springer A, Mutran E, Manson SM, et al. Ethnic minority older adults participating in clinical research: developing trust. J Aging Health 2004 Nov;16(5 Suppl):93S-123S. [doi: 10.1177/0898264304268151] [Medline: 15448289]
10. Cho J, Choi E, Kim SY, Shin DW, Cho B, Kim C, et al. Association between cancer stigma and depression among cancer survivors: a nationwide survey in Korea. Psychooncology 2013 Oct;22(10):2372-2378. [doi: 10.1002/pon.3302] [Medline: 23784964]

11. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: Systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
12. Statista. 2015. Facebook - Statistics & Facts URL: https://www.statista.com/topics/751/facebook/ [WebCite Cache ID 6yngKnZCq]
13. York A. Sprout Social. Social Media Demographics to Inform a Better Segmentation Strategy URL: http://sproutsocial.com/insights/new-social-media-demographics/ [WebCite Cache ID 6yngceeiz]
14. Greenwood S, Perrin A, Duggan M. Social media update. 2016 Nov 11. Social Media Update 2016: Facebook usage and engagement is on the rise, while adoption of other platforms holds steady URL: http://www.pewinternet.org/2016/11/11/social [WebCite Cache ID 6yiPwOV4u]
15. Anderson M, Perrin A. Pew Research Center. 2017 May 17. Tech Adoption Climbs Among Older Adults: Technology use among seniors URL: http://www.pewinternet.org/2017/05/17/technology-use-among-seniors/ [WebCite Cache ID 6yngceeiz]
16. Lu Q, Gallagher MW, Loh A, Young L. Expressive Writing Intervention Improves Quality of Life Among Chinese-American Breast Cancer Survivors: A Randomized Controlled Trial. Ann Behav Med 2018 Oct 22;52(11):952-962. [doi: 10.1093/abm/kax067] [Medline: 30346497]
17. Kearney N, McCann L, Norrie J, Taylor L, Gray P, McGee-Lennon M, et al. Evaluation of a mobile phone-based, advanced symptom management system (ASyMS) in the management of chemotherapy-related toxicity. Support Care Cancer 2009 Apr;17(4):437-444. [doi: 10.1007/s00520-008-0515-0] [Medline: 18953579]
18. Whitaker C, Stevelink S, Fear N. The Use of Facebook in Recruiting Participants for Health Research Purposes: A Systematic Review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
19. Schumacher KR, Stringer KA, Donohue JE, Yu S, Shaver A, Caruthers RL, et al. Social media methods for studying rare diseases. Pediatrics 2014 May;133(5):e1345-e1353 [FREE Full text] [doi: 10.1542/peds.2013-2966] [Medline: 24733869]

Edited by G Eysenbach; submitted 12.07.18; peer-reviewed by L Bruzzone, L Nguyen; comments to author 17.10.18; revised version received 17.10.18; accepted 23.10.18; published 10.01.19.

Please cite as:
Tsai W, Zavala D, Gomez S
Using the Facebook Advertisement Platform to Recruit Chinese, Korean, and Latinx Cancer Survivors for Psychosocial Research: Web-Based Survey Study
J Med Internet Res 2019;21(1):e11571
URL: http://www.jmir.org/2019/1/e11571/
doi:10.2196/11571
PMID:30632966

Original Paper

How Search Engine Data Enhance the Understanding of Determinants of Suicide in India and Inform Prevention: Observational Study

Natalia Adler[1], MA; Ciro Cattuto[2], PhD; Kyriaki Kalimeri[2], PhD; Daniela Paolotti[2], PhD; Michele Tizzoni[2], PhD; Stefaan Verhulst[3], MA; Elad Yom-Tov[4], PhD; Andrew Young[3], MA

[1]United Nations International Children's Emergency Fund (UNICEF), New York, NY, United States

[2]ISI Foundation, Torino, Italy

[3]The Governance Lab, New York University, New York, NY, United States

[4]Microsoft Research, Herzeliya, Israel

Corresponding Author:
Daniela Paolotti, PhD
ISI Foundation
Via Chisola 5
Torino, 10126
Italy
Phone: 39 011 660 3090
Email: daniela.paolotti@isi.it

Abstract

Background: India is home to 20% of the world's suicide deaths. Although statistics regarding suicide in India are distressingly high, data and cultural issues likely contribute to a widespread underreporting of the problem. Social stigma and only recent decriminalization of suicide are among the factors hampering official agencies' collection and reporting of suicide rates.

Objective: As the product of a data collaborative, this paper leverages private-sector search engine data toward gaining a fuller, more accurate picture of the suicide issue among young people in India. By combining official statistics on suicide with data generated through search queries, this paper seeks to: add an additional layer of information to more accurately represent the magnitude of the problem, determine whether search query data can serve as an effective proxy for factors contributing to suicide that are not represented in traditional datasets, and consider how data collaboratives built on search query data could inform future suicide prevention efforts in India and beyond.

Methods: We combined official statistics on demographic information with data generated through search queries from Bing to gain insight into suicide rates per state in India as reported by the National Crimes Record Bureau of India. We extracted English language queries on "suicide," "depression," "hanging," "pesticide," and "poison". We also collected data on demographic information at the state level in India, including urbanization, growth rate, sex ratio, internet penetration, and population. We modeled the suicide rate per state as a function of the queries on each of the 5 topics considered as linear independent variables. A second model was built by integrating the demographic information as additional linear independent variables.

Results: Results of the first model fit (R^2) when modeling the suicide rates from the fraction of queries in each of the 5 topics, as well as the fraction of all suicide methods, show a correlation of about 0.5. This increases significantly with the removal of 3 outliers and improves slightly when 5 outliers are removed. Results for the second model fit using both query and demographic data show that for all categories, if no outliers are removed, demographic data can model suicide rates better than query data. However, when 3 outliers are removed, query data about pesticides or poisons improves the model over using demographic data.

Conclusions: In this work, we used search data and demographics to model suicide rates. In this way, search data serve as a proxy for unmeasured (hidden) factors corresponding to suicide rates. Moreover, our procedure for outlier rejection serves to single out states where the suicide rates have substantially different correlations with demographic factors and query rates.

(J Med Internet Res 2019;21(1):e10179) doi:10.2196/10179

KEYWORDS

internet data; India; suicide; mobile phone

Introduction

Background

According to the World Health Organization (WHO), close to 800,000 people die by suicide every year, with 78% of global suicides occurring in low- and middle-income countries [1]. Teenagers and young adolescents are particularly at risk, as suicide represents the second leading cause of death among 15-29-year-olds worldwide [1]. These concerning figures do not even fully capture the magnitude of the problem. The WHO estimates that good quality data on suicide exist for only 60 countries worldwide.

According to official statistics, India is home to 20% of the world's suicide deaths [2], yet the issue attracts limited national public health attention [3]. In addition, statistics on suicides released by the Indian National Crime Records Bureau (NCRB) are insufficient to understand the magnitude of the problem. In 2013, the NCRB reported that 134,799 people died of suicide, making the suicide rate 11% of total deaths [4]. However, evidence from other studies shows that the NCRB's suicide rate data are grossly underreported. For instance, the WHO reported 170,000 cases of suicide deaths in India, which is about 35,000 higher than the NCRB's data [3]. Similarly, the Registrar General of India implemented a nationally representative mortality survey indicating that about 3% of the surveyed deaths (2684 of 95,335) in individuals aged 15 years or older were due to suicide, corresponding to about 187,000 suicide deaths in India in 2010 (as reported in Patel et al [3]).

The factors contributing to these inconsistencies are likely manifold, including both data collection barriers and cultural challenges. The deep-rooted stigma associated with mental disorders, coupled with limited suicide prevention and mental health services, makes it difficult to address suicide as a major public health problem in India [5]. Until recently, suicide was a criminal offense in the country, likely compelling families to report suicides as death by an illness or accident so as to avoid punishment [6]. Moreover, analysis (if any) of suicide records is limited to demographic correlations. Patel et al [3], for instance, only focused on age and sex variables to analyze the survey findings.

Furthermore, there is little research on the important role played by stigma in suicide reporting in India, with the majority of the studies stressing the necessity of further research for a systematic assessment. Some existing studies, even if they do not provide a systematic assessment of the topic, report a connection between suicide reporting and stigma. Merriott [7], in a study of factors associated with the farmer suicide crisis in India, acknowledges the presence of stigma associated with suicide underreporting: "The NCRB figures, for which the studies in the introduction proposing an increasing farmer suicide rate come from, are considered significant underestimates as, for example, they only use police records to classify deaths, and due to the stigma associated with suicide in a country where it was illegal until a government decision in 2014." Bhise and Behere [8] stress the presence of stigma related to mental illnesses and suicide prevention without proceeding to an assessment of stigma: "Creating a referral network of government and private hospitals and mental health professionals, training health professionals in identifying high-risk farmers, and strategies aimed at reducing stigma attached to mental illness will go a long way in suicide prevention." Similarly, the study by Aggarwal [4]refers to the stigma related to suicidal behavior: "The anticipated changes as a result of this policy shift include: accurate reporting and recording of suicide as a cause of death, reduction in stigma associated with suicidal behavior and use of these figures to inform suicide prevention strategies." A case study in Pakistan by Kahn et al [9] states that "the absence of systematic sampling of police data in societies with high social stigma will oversample people with severe mental illness, suggests selection bias and probably invalidates the results." Kennedy et al [10] demonstrated higher levels of stigma and higher levels of suicide literacy in a study conducted in the Australian rural farming communities, suggesting how best practices can be adapted to improve stigma reduction and suicide prevention efforts.

Finally, a study by Armstrong and collaborators [11] focuses on how media reporting of suicide news in India performs against the WHO guidelines, stressing how strategies should be devised to boost the positive contribution that media can make to suicide reporting and prevention.

This paper seeks to contribute to this discussion by describing how data gaps in Indian suicide reporting can be filled through the creation of data collaboratives. Data collaboratives are "an emerging form of public-private partnership in which actors from different sectors exchange information to create new public value." [12]. Data collaboratives are increasingly being tested as a means for improving evidence-based policy making and targeted service delivery around the world, including, notably, data-sharing arrangements between corporations and national statistical offices [13].

The main goal of this paper is to shed some light on the public health issues of suicides among the Indian population through the lens of Web data. Can data about Web-based information-seeking behavior help to study the determinants for suicides in the various Indian states? In particular, by combining official statistics on suicide with demographic information about the population and data generated through search queries on various keywords related to suicide, this paper seeks to: add an additional layer of information to more accurately represent the magnitude of the problem, determine whether search query data can serve as an effective proxy for studying factors contributing to suicide that are not represented in traditional datasets (eg, search queries for specific keywords related to means of suicide or to social factors that can influence the mental status of the person, such as economic difficulties or academic pressure for young people), and consider how data collaboratives built on search query data could inform future suicide prevention efforts in India and beyond.

What Is Known About Those Who Die by Suicide in India

The literature varies and sometimes offers a convoluted picture of who is at risk of dying by suicide, both in India and beyond. There seems to be a consensus that young people aged 15-29 years are particularly at risk. Patel et al [3] found that 40% of

suicides among men and 56% of suicides among women occurred between the ages of 15 and 29 years. However, while the prevalence of suicide among the young is generally accepted, the male-to-female suicide ratio in India varies greatly, ranging from 1.04 to 1.63 according to some studies [6], while official statistics from the NCRB show a male-to-female suicide ratio of 2:1. Patel et al [3], on the other hand, found an age-standardized rate per 100,000 people aged 15 years or older of 26.3 for men and 17.5 for women, demonstrating the need for more clarity. The paper also found that Indian boys and men had a 1.7% cumulative risk of dying by suicide between the ages of 15 and 80 years compared with a 1.0% risk among girls and women [4].

There exists a strong correlation between educational backgrounds and suicide risks, with the less educated accounting for 70.4% of suicide cases as recorded in the NCRB data [6]. Counterintuitively, suicide among students in India is also increasing, moving from 5.5% of total suicides in 2010 to 6.2% in 2013 [14].

Suicide rates vary by occupation [4]. Housewives accounted for about 18% of the total victims, while farmers comprised 11.9% of the total victims followed by those working in the private sector (7.8%), unemployed (7.5%), and those working in the public sector (7.8% and 2.2%, respectively). A study reported that approximately 16,000 farmers in India die by suicide every year [7]. Patel et al [3] found that about half of suicide deaths in India arose from poisoning, especially resulting from the ingestion of pesticides.

Risk Factors

Some population groups are more at risk than others. For example, a study shows that 27.2% of primary care patients suffer from depressive disorder, and 21.3% of them have attempted suicide, demonstrating how depression is 1 of the underlying factors that drives suicide [15].

Risk factors cut across geographical lines, with official statistics showing a significantly higher rate of suicide taking place in southern states, such as Tamil Nadu, Andhra Pradesh, Karnataka, Kerala, West Bengal, and Maharashtra, where 63.6% of suicide cases occurred [4]. South India is the area encompassing the Indian states of Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, and Telangana as well as the union territories of Lakshadweep, Andaman and Nicobar Islands, and Puducherry, occupying 19% of India's area and with about 18% of the total population of India.

These data paint a picture of the breadth and diversity of the suicide issue in India. But given the stigma associated with suicide, poor quality data, and the still-recent decriminalization of suicide attempts, these statistics confuse as much as they elucidate. The country's underreporting challenge—and the likely neglect of certain population groups altogether—creates major challenges for meaningfully determining who is at risk.

Internet Data as a Source for Health Information

In response to the issue of suicide underreporting in India, this paper looks at a specific cohort of people *(* English-speaking internet users*)* to add another layer of understanding about those at risk of suicide. The majority (one-third) of internet users are young (18-35 years) [16], which coincides with the age group most at risk for suicides (15-29 years).

With its 360 million users (26% of the population), India is home to more internet users than any country save China [17]. Men dominate internet usage in India [18], with 71% to women's 29%. Internet usage is more prevalent in northern (27%) and western states (25%) compared with the South (19%) and East (16%) [16]. Yet, not all Indians accessing the internet do so in English. The Indian Constitution recognizes 22 official languages, and the number of Indian-language internet users has grown dramatically over the years, surpassing English users: 234 million compared with 175 million, respectively.

Internet data have been used to monitor health behaviors for a variety of conditions ranging from infectious diseases [19-21] to mental health conditions [22]. Social media is one source of internet data that has provided several insights on suicides [23-25]. However, as opposed to social media, anonymous Web-based venues, especially search engines [26], allow people to seek information on sensitive topics. Monitoring such venues can, thus, offer a window into behaviors that are otherwise difficult to study. Specifically, in the case of suicides, social media has been used to detect suicidality [27], and search engine logs were utilized to analyze suicides in general [28-30] and the Werther Effect (copycat suicides) in particular [21]. Additionally, search volume for past versus future was shown to be a predictor of suicide rates in the United States [31].

Here we examine internet search engine logs for information about suicides. Search engine logs, as analyzed here, focus on a population of English-speaking people in India. The market share of Bing in India was reported to be around 7% at the time of data collection [32]. Moreover, as shown in Fisher and Yom-Tov [33], people seeking information on suicides via search engines are (at least in the United States) people who are contemplating suicide, not people who may necessarily die by suicide.

Internet Search Queries as a Source for Suicide-Related Information

This paper's methodology, described below, builds on previous work leveraging search query data analysis. Numerous studies have found that search query data are reflective of behaviors in the physical world [34]. In the United States, for example, people searching for actionable information about suicides (how to kill themselves) correspond to the population that attempts suicide—but not the population that successfully suicides [33].

Several studies have analyzed Google Trends, an aggregate measure of search query volume, and found correlations between search queries for suicide and the rate of suicide. Gunn and Lester [28] found a correlation between the volume of queries about suicide and the actual number of suicides by analyzing search words and phrases like "how to suicide." Hagihara et al [29] conducted a study in Japan that shows how suicide queries spike in the period before there is an increase in suicide rate [24]. This method was replicated in Taiwan and Australia, but those studies yielded contradicting results [28]. Other studies are most skeptical about the correlation between Google queries

and suicide rates, concluding that a tool to identify relevant search queries must be further developed to create a more precise modeling mechanism [35].

Finally, Kristoufek et al [36] studied how data on the number of Google searches for the terms "depression" and "suicide" in England related to the number of suicides between 2004 and 2013. The researchers found that estimates drawing on Google data were significantly more accurate than estimates relying on previous suicide data alone. Interestingly, their findings show that a greater number of searches for the term "depression" is related to fewer suicides, whereas a greater number of searches for the term "suicide" is related to more suicides, though the correlation is not extremely high (R^2 of about 0.4).

Methods

Search Engine Data

We extracted all English language queries from the Bing search engine submitted by people from India between November 2016 and February 2017 (inclusive). For each query, we recorded the time and date of the query, the state in India from where the user made the query, and the text of the query. The correlation between the number of queries per state and the population of that state multiplied by internet penetration provided a positive Spearman correlation (ρ=0.93, P<.001).

The queries on 5 topics were identified by testing whether the text of the queries contained 1 or more of the inclusion terms in Table 1 and did not contain any of the exclusion terms. The exclusion terms were found by identifying the most common words and word pairs appearing in conjunction with the inclusion terms and identifying those that were unrelated to suicidal intentions.

State Data on Suicide Rates

The suicide rate per state was obtained from the latest available data, the 2014 Accidental Deaths & Suicides in India report from the NCRB [37] (see Multimedia Appendix 1).

Demographic Data

In the data analysis, we have included demographic information at the state level, including urbanization, growth rate, sex ratio, internet penetration, and population. Data were obtained as follows:

1. Sex ratio, population, urbanization, and growth rate: from Wikipedia [38].
2. Income: per capita national income 2013-2014, available from the India National Informatics Centre [39].
3. Internet penetration: from data published by The Hindu newspaper [40].
4. Enrollment in higher education: gross enrollment ratio in higher education, available from the Statistical Year Book of India, 2016 [41].

Statistical Modeling of Search Engine Data

Data were analyzed for their temporal patterns (diurnal and weekly) as well as their variation by state. We modeled the reported suicide rate per state as a function of the fraction of queries on each of the 5 topics from each state. Thus, the dependent variable in our models was the expected number of suicides in each state, which is the product of the reported suicide rate multiplied by the size of the population. The independent variables included the fraction of queries (with respect to the number of internet users in the state) from each state for each topic. Outliers were removed using an iterative process: up to either 3 or 5 states were removed by finding the state which, if removed, increased model fit (R^2) by the greatest amount and repeating this process 3 or 5 times, as desired. The model used throughout is a linear model, unless otherwise stated. We report model fit for different levels of outlier rejection below.

Risks and Data Responsibility

To be clear, the use of data related to suicide, suicidal ideation, and mental health creates some level of risk across the data lifecycle. The analysis described in this paper adhered to strict data responsibility principles, ensuring that sensitive data were not shared or compromised and that aggregated rather than personally identifiable data informed our findings.

For the field at large, to effectively and legitimately leverage data collaboratives to improve public understanding of suicide rates and devise evidence-based prevention strategies, data responsibility methods and tools are needed for both sides of data-sharing arrangements.

Table 1. Exclusion and inclusion terms for each of the 5 topics related to suicides.

Topic	Inclusion terms	Exclusion terms
Suicide	"suicide," "kill myself"	"suicide squad," "song," "download," "skill," "killer," "movie," "video," "bill," "game," "lyrics," "mp3," "suicide girl," "militia," "mockingbird," "ghandi," "akame ga kill," "3 days to kill," "wifi kill," "kill dil," "kill zone," "killzone," "kill em with kindness," "kill me heal me," "rkill"
Depression	"depression," "depressed"	—[a]
Hanging	"hang," "hanging"	"wall hanging," "hanging garden," "macrame"
Pesticide	"pesticide"	—
Poison	"poison"	"poison ivy," "poisonous snakes," "poisoned thoughts," "poison thoughts," "food poisoning," "hanging boobs," "hanging lights"

[a]No exclusion terms considered.

The research that informed this paper is contributing to the development of data responsibility frameworks to aid the field in assessing if, when, and how data can be shared in a responsible manner as part of a data collaborative. This study is considered exempt by the Microsoft Institutional Review Board.

Results

Temporal Analysis

Figure 1 shows the percentage of queries about suicide, depression, and suicide methods as a function of the hour of the day and the day of the week. Days are numbered sequentially from Sunday (1) through Saturday (7). As these figures show, these queries broadly follow the baseline (all queries made in India). However, closer inspection reveals that relevant queries are approximately 20% less likely, compared with the baseline, during early morning hours and up to 15% more likely during the evening to late night hours. The largest difference between the baseline and relevant queries when stratified by day of the week is smaller than 5%.

Using Query Data to Model Suicide Rate

Table 2 shows the model fit (R^2) when determining the suicide rates from the fraction of queries in each of the 5 topics as well as the fraction of all suicide methods. States with <0.25% of the Indian population are excluded (n=20). As the table shows, the correlation increases significantly with the removal of even 3 outliers and improves slightly when 5 outliers are removed. In all cases, a statistically significant correlation is reached, but the best correlation is obtained for suicide methods (hanging, pesticide, and poison) and only to a lesser extent for depression. This indicates that people who are considering suicide are not only just asking about the term itself but also about possible precursors (depression) and methods of suicide.

We next analyzed the outliers and whether, given a model that was constructed after the exclusion of an outlier, more suicides would be modeled by query volume compared with official statistics, or vice versa. A negative outlier would, thus, indicate that more suicides are determined by the model according to the volume of queries in a topic than are reported by official data. A positive outlier would indicate the reverse: the official reported suicide rate is greater than that which would be inferred from the queries.

Analyzing the models after excluding 5 outliers per model, we find that there are slightly more negative outliers than positive ones: 16 negative outliers compared with 14 positive outliers.

Figure 1. Diurnal and weekly patterns of relevant queries (suicide, depression, and suicide methods) compared to the baseline of all queries made in India.

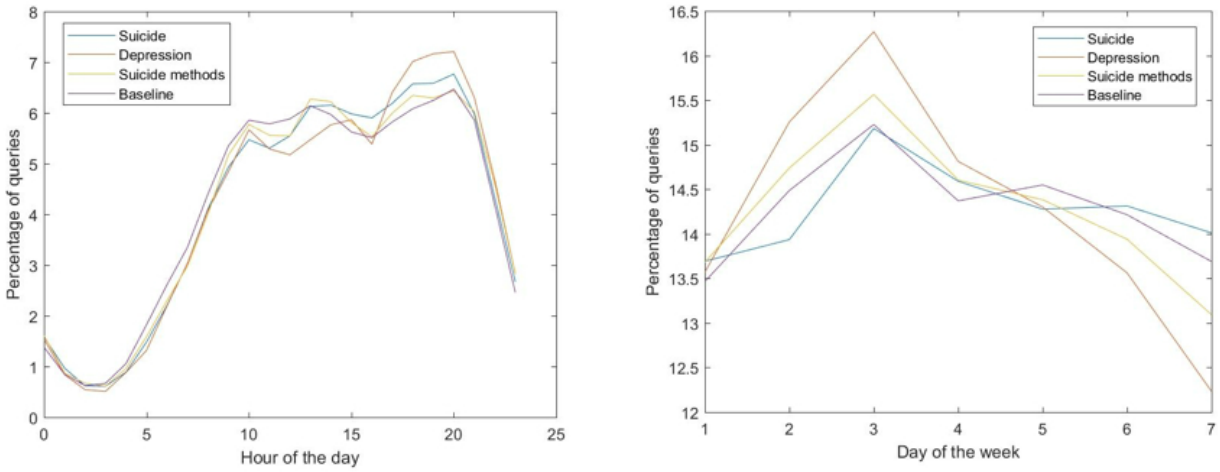


Table 2. Model fit for modeling the expected number of suicides in each state from the fraction of queries in each topic.

Query	R^{2a} (no outliers removed)	R^2 after removal of 3 outliers	R^2 after removal of 5 outliers
Hanging	0.29	0.49	0.65
Pesticide	0.16	0.71	0.80
Poison	0.33	0.65	0.72
All methods of suicide (hanging, pesticide, and poison)	0.47	0.68	0.80
Suicide	0.13	0.65	0.79
Depression	−0.01	0.34	0.50

[a]R^2: model fit.

Figure 2 shows the outliers. From this figure, it can be seen that the outliers are not distributed randomly; in all but 1 case, a state will either have all positive or all negative outliers. The states with most negative outliers are Jammu & Kashmir and Jharkhand (4 and 3 outliers, respectively), and the ones with most positive outliers are Telangana and Gujarat (5 and 4 outliers, respectively).

Our findings singled out 2 states, Jammu & Kashmir and Jharkhand, as having more queries indicative of suicide than would be expected, given the published suicide rates in these states. In contrast, Gujarat and Telangana have more reported suicides than modeled by search data.

Addition of Demographic Data to Query Data

Demographic data are correlated with suicide rate. Therefore, in this section, we investigate whether query data can add to the modeling of suicide rates, beyond what demographic data can provide. To overcome the dimensionality of additional variables, we employ a stepwise linear model, which selects the most significant variables under the criteria that the *P* value for an *F* test of the change in the sum of squared error is maximally reduced by adding a variable.

Figure 2. Maps showing states that are negative outliers (more suicides modeled by Web data than reported), left, and positive outliers (fewer suicides modeled by Web data than reported), right. Darker colors indicate that the state is an outlier in more terms.

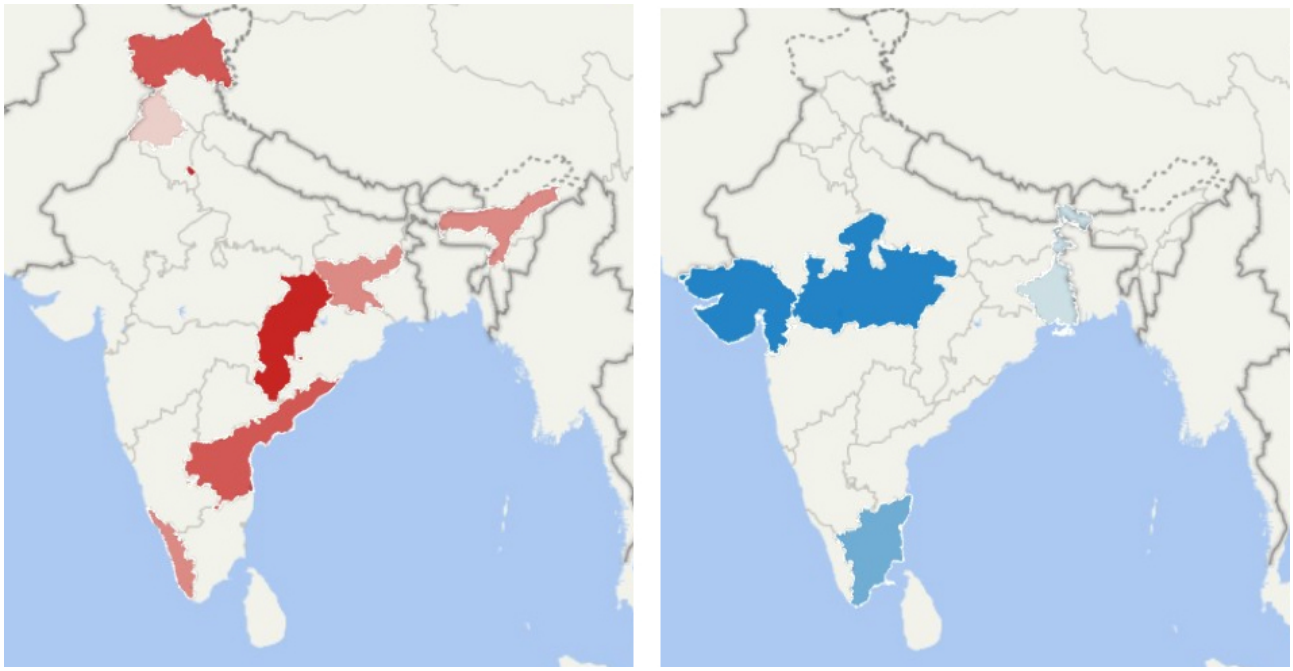


Table 3. Model fit for the expected number of suicides in each state from the fraction of queries in each topic, with and without demographics, using a stepwise model.

Query	Without demographic data		With demographic data	
	R^2[a] (no outliers removed)	R^2 after removal of 3 outliers	R^2 (no outliers removed)	R^2 after removal of 3 outliers
Hanging	0.28	0.49	0.51[b]	0.75[b]
Pesticide	0.16	0.71	0.51[b]	0.91
Poison	0.33	0.65	0.51	0.76
All methods of suicide (hanging, pesticide, and poison)	0.47	0.68	0.47	0.74
Suicide	0.13	0.34	0.51[b]	0.75[b]
Depression	−0.01	0.65	0.51[b]	0.75[b]

[a]R^2: model fit.

[b]Cases where query data were not selected for inclusion in the model.

Table 4. Outliers (with the rejection of 3 states; the direction of outliers is in parentheses) for pesticides and poison queries for models that use demographic data and query data and for models that only use query data. + and – indicate positive and negative directions, respectively.

Query	Demographics + query data model	Query data only
Pesticides		
Andhra Pradesh	–	–
Kerala	–	N/A[a]
Punjab	–	–
Jammu and Kashmir	N/A	–
Poison		
Telangana	+	+
Delhi	+	N/A
Jharkhand	–	–
Madhya Pradesh	N/A	+

[a]N/A: not applicable.

Table 3 shows the model fit for models using solely query data and for models using both query data and demographic data. States with less than 0.25% of the Indian population are excluded (n=20)The variables selected when using just demographic data are urbanization rate and sex ratio, both positively correlated with suicide rates. Table 3 shows that for all categories, if no outliers are removed, demographic data can model suicide rates better than query data. However, when 3 outliers are removed, query data about pesticides or poisons becomes significant (fourth column) and improves the model over using only demographic data.

Using both demographics and query data, the outlying states for pesticides are all negative (more suicides modeled by Web data than reported), namely, Andhra Pradesh, Kerala, and Punjab. The outliers for poison are Telangana (positive), Delhi (negative), and Jharkhand (negative). Thus, the sole positive outlier is the same as using query data alone.

Comparing the outliers we obtained with the query data only with those obtained when including the demographics data, results are rather similar (Table 4). We report only outliers for pesticides and poison because these are the only keywords for which query data were selected for inclusion in the model.

Discussion

Internet search data have been shown in previous work to serve as a proxy for many health-related behaviors, enabling the measurement of rates of different conditions ranging from influenza to suicide. Here we use these data to model suicide rates in India. Internet data can be influenced and biased by the population using it. Similarly, official suicide statistics might be susceptible to underreporting by statistical agencies. In our work, we, therefore, applied 2 tools that could mitigate these biases. First, we used both search data and demographics to enhance the understanding of official suicide rates. In this way, search data serve as proxy for unmeasured (hidden) factors corresponding to suicide rates. Second, our procedure for outlier rejection serves to single out states where the suicide rates have substantially different correlations with both demographic factors and query rates. To emphasize the difference between these 2 influences, consider the following simplifying cases as clarifying examples: First, suppose that in 1 state there exists a population where people kill themselves by suicide but does not use the internet to search for it beforehand. In that case, such a state would be identified as an outlier in our data. Similar cases would occur if there is underreporting in the state or some demographic factor not included in the model that does not influence the search queries. In this case, the state would be highlighted as a positive outlier. On the other hand, if a demographic factor that is unavailable to the model can help understand suicide rates and search queries are influenced by this factor, they will serve as a proxy to suicide rates and will improve model correlations. We believe that in the data analyzed, a mix of these effects is at work. Specifically, some (mostly agricultural states) were found to be negative outliers; in these states, it might be that a population who does not use the internet or the Bing search engine are those among which more suicides are reported. Similarly, several states were identified as positive outliers, suggesting that in those states, underreporting might be occurring, or there might be some other social or demographic factor at play that is not captured by the model and by the search queries activity. We do not know exactly which of these effects are at play. Thus, further investigation will be needed to disentangle social factors from actual suicide underreporting.

Cases in point are Telangana and Andhra Pradesh. The former was recently separated from the latter and declared an independent state. Both are states where agriculture is a major industry and, thus, where farmer suicides may be expected. However, the first of these was identified as a positive outlier (where fewer suicides were modeled based on query data) whereas the latter was identified as the opposite, for both models. Several explanations may be offered for this difference. First, Telangana has registered more suicides due to poverty and unemployment [37]. Second, Telangana has a higher urbanization rate compared with Andhra Pradesh (39% and 30%, respectively). Third, Telangana has a higher education rate (35% vs 29%). Both variables are taken into account by

the demographics and could, thus, show a higher suicide rate in Telangana than is expected solely from query data. This is coupled with the fact that Telangana has recorded a recent increase in student suicides [42]. Finally, in an attempt to curb farmer suicides, Digital India has improved access to information by providing farmers with internet access [43], a factor which may have contributed to a higher than expected rate of queries from farmers in our data.

Regarding methods of suicide, queries for 2 methods were found to improve the modeling of suicide rate over demographics. Interestingly, searches for depression or the phrase "suicide" did not. This result has an important implication: efforts to model the incidence of suicide toward taking preventative action are more likely to find success if they focus on queries about specific methods of suicide as opposed to keywords related to depressive symptoms or the concept of suicide more generally.

Most internet search engines nowadays provide information on helpline numbers in highlighted information boxes above search results when users search for information on suicides. Such pointers to crisis support, particularly in smartphone apps, have been shown to be effective in some studies [44]. Our data on the diurnal variations and weekly variations of queries can help guide the staffing of suicide prevention helplines. Moreover, to the best of our knowledge, these boxes are only displayed when people explicitly search for the term "suicide." Our results suggest that future research should investigate the display of these boxes also in cases where people search for methods of suicide. However, it is unclear how to distinguish between searches of methods that are related to suicides and those that are not (eg, a farmer searching for information on pesticides).

Public awareness campaigns may be a driver of people searching for information about suicides. However, in the data analyzed, we found that searches for methods of suicide are better correlated with the suicide rate. We suggest that this shows that our data are not affected by such exogenous factors. We also suggest that additional research is needed to explore the feasibility of creating new recommendations on the sale of pesticides in India, particularly for sales to young people.

Building on this work and drawing upon cross-sector data, including but not limited to search queries, we are conducting research aimed at increasing our understanding of the drivers of suicide among young people in India, how those drivers differ across regions, and how those findings can inform suicide prevention efforts in the country going forward.

Acknowledgments

All the authors of the work would like to express their gratitude to Brinda Adige, Global Concerns India; Trisha Shetty, SheSays; and Purva Sawant at United Nations International Children's Emergency Fund for their insightful input and their valuable comments and suggestions. Thanks also to Michelle Winowatan, The Governance Lab, for important research support.

DP, MT, CC, and KK acknowledge partial support from the Lagrange Project of ISI Foundation funded by the CRT (Cassa di Risparmio di Torino) Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary information on the data sources.

[PDF File (Adobe PDF File), 67KB - jmir_v20i12e10179_app1.pdf]

References

1. World Health Organization. 2018 Aug 24. Suicide URL: http://www.who.int/en/news-room/fact-sheets/detail/suicide [WebCite Cache ID 6xNNftr6f]
2. World Health Organization. 2017 Mar. Suicide rate estimates, age-standardized estimates by country URL: http://apps.who.int/gho/data/view.main.MHSUICIDEASDRv?lang=en [accessed 2018-02-20] [WebCite Cache ID 6xNIXT2mc]
3. Patel V, Ramasundarahettige C, Vijayakumar L, Gajalakshmi V, Gururaj G, Suraweera W, et al. Million Death Study Collaborators, Suicide mortality in India: a nationally representative survey. The Lancet 2012 Jun;379(9834):2343-2351. [doi: 10.1016/S0140-6736(12)60606-0] [Medline: 22726517]
4. Aggarwal S. Suicide in India. Br Med Bull 2015 Jun;114(1):127-134. [doi: 10.1093/bmb/ldv018] [Medline: 25958380]
5. World Health Organization. 2012. Public Health Action for the Prevention of Suicide: A Framework URL: http://www.webcitation.org/6xNJsLc9B [WebCite Cache ID 6xNJsLc9B]
6. Radhakrishnan R, Andrade C. Suicide: An Indian perspective. Indian Journal of Psychiatry 2012;54(4):304 [FREE Full text]
7. Merriott D. Factors associated with the farmer suicide crisis in India. J Epidemiol Glob Health 2016 Dec;6(4):217-227 [FREE Full text] [doi: 10.1016/j.jegh.2016.03.003] [Medline: 27080191]
8. Bhise M, Behere P. Risk factors for farmers' suicides in central rural India: Matched case-control psychological autopsy study. Indian Journal of Psychological Medicine 2016;38(6):560. [doi: 10.4103/0253-7176.194905]

9. Khan MM, Mahmud S, Karim MS, Zaman M, Prince M. Case-control study of suicide in Karachi, Pakistan. Br J Psychiatry 2008 Nov;193(5):402-405. [doi: 10.1192/bjp.bp.107.042069] [Medline: 18978322]
10. Kennedy AJ, Brumby SA, Versace VL, Brumby-Rendell T. Online assessment of suicide stigma, literacy and effect in Australia's rural farming community. BMC Public Health 2018 Jul 06;18(1):846 [FREE Full text] [doi: 10.1186/s12889-018-5750-9] [Medline: 29980237]
11. Armstrong G, Vijayakumar L, Niederkrotenthaler T, Jayaseelan M, Kannan R, Pirkis J, et al. Assessing the quality of media reporting of suicide news in India against World Health Organization guidelines: A content analysis study of nine major newspapers in Tamil Nadu. Aust N Z J Psychiatry 2018 Sep;52(9):856-863. [doi: 10.1177/0004867418772343] [Medline: 29726275]
12. Susha I, Janssen M, Verhulst S. Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development. Open Data, Information Processing, and Datification in Government 2017 Jan:2691 [FREE Full text] [doi: 10.24251/HICSS.2017.325]
13. Klein T, Verhulst S. Access to new data sources for statistics: Business models and incentives for the corporate sector. OECD Statistics Working Papers 2017;06:2. [doi: 10.1787/9a1fa77f-en]
14. Ponnudurai R. Suicide in India: changing trends and challenges ahead. Indian Journal of Psychiatry 2015;57(4):348 [FREE Full text]
15. Indu P, Anilkumar T, Pisharody R, Russell P, Raju D, Sarma P, et al. Prevalence of depression and past suicide attempt in primary care. Asian J Psychiatr 2017 Jun;27:48-52. [doi: 10.1016/j.ajp.2017.02.008] [Medline: 28558895]
16. Pew Research Center. 2015. Profile of Indian internet users URL: http://www.pewresearch.org/fact-tank/2016/04/06/global-tech-companies-see-indias-vast-offline-population-as-untapped-market/ft_16-04-06_indiainternet_users/ [accessed 2018-02-20] [WebCite Cache ID 6xNO2R2P9]
17. Wikipedia. 2016. List of countries by number of Internet users URL: https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users [WebCite Cache ID 6xNLfoK3J]
18. Statista. 2015. Distribution of internet users in India as of October 2015, by gender URL: https://www.statista.com/statistics/272438/gender-distribution-of-internet-users-in-india/ [WebCite Cache ID 6xNLmXt56]
19. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. 2006 Presented at: AMIA Annual Symposium Proceedings; November 11-15, 2006; Washington, DC.
20. Polgreen P, Chen Y, Pennock D, Nelson F. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 01;47(11):1443-1448. [doi: 10.1086/593098] [Medline: 18954267]
21. Ginsberg J, Mohebbi M, Patel R, Brammer L, Smolinski M, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
22. Reece A, Reagan A, Lix K, Dodds P, Danforth C, Langer E. Forecasting the onset and course of mental illness with Twitter data. Scientific Reports 2017 Dec;7(1):13006 [FREE Full text]
23. Jashinsky K, Burton S, Hanson C, West J, Giraud-Carrier J, Barnes M, et al. Tracking Suicide Risk Factors Through Twitter in the US. Crisis 2014 Jan;35(1):51-59 [FREE Full text] [doi: 10.1027/0227-5910/a000234]
24. O'Dea B, Wan S, Batterham P, Calear A, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interventions 2015 May;2(2):183-188. [doi: 10.1016/j.invent.2015.03.005]
25. Zhang L, Huang X, Liu T, Li A, Chen Z, Zhu T. Using Linguistic Features to Estimate Suicide Probability of Chinese Microblog Users. In: Human Centered Computing. 2014 Presented at: International Conference on Human Centered Computing; November 27-29 2014; Phnom Penh, Cambodia p. 549-559 URL: https://link.springer.com/chapter/10.1007/978-3-319-15554-8_45
26. Pelleg D, Yom-Tov E, Maarek Y. Can You Believe an Anonymous Contributor? On Truthfulness in Yahoo! Answers. 2012 Presented at: Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom); September 3-5, 2012; Amsterdam, The Netherlands p. 420. [doi: 10.1109/SocialCom-PASSAT.2012.13]
27. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In: Proc SIGCHI Conf Hum Factor Comput Syst. 2016 May Presented at: CHI '12 CHI Conference on Human Factors in Computing Systems; May 5-10, 2012; Austin, Texas, USA p. 2098-2110 URL: http://europepmc.org/abstract/MED/29082385 [doi: 10.1145/2858036.2858207]
28. Gunn JFIII, Lester D. Using google searches on the internet to monitor suicidal behavior. Journal of Affective Disorders 2013 Jun;148(2-3):411-412. [doi: 10.1016/j.jad.2012.11.004]
29. Hagihara A, Miyazaki S, Abe T. Internet suicide searches and the incidence of suicide in young people in Japan. Eur Arch Psychiatry Clin Neurosci 2012 Feb;262(1):39-46. [doi: 10.1007/s00406-011-0212-8] [Medline: 21505949]
30. Tana J, Kettunen J, Eirola E, Paakkonen H. Diurnal Variations of Depression-Related Health Information Seeking: Case Study in Finland Using Google Trends Data. JMIR Ment Health 2018 May 23;5(2):e43 [FREE Full text] [doi: 10.2196/mental.9152] [Medline: 29792291]
31. Lee D, Lee H, Choi M. Examining the Relationship Between Past Orientation and US Suicide Rates: An Analysis Using Big Data-Driven Google Search Queries. J Med Internet Res 2016 Feb 11;18(2):e35 [FREE Full text] [doi: 10.2196/jmir.4981] [Medline: 26868917]

32. Statista. 2017. Worldwide search market share of Bing as of August 2017, by country URL: http://www.webcitation.org/6xNM1Y616 [WebCite Cache ID 6xNM1Y616]
33. Yom-Tov E, Fischer S. The Werther Effect Revisited: Measuring the Effect of News Items on User Behavior. In: Proceedings of the 26th International Conference on World Wide Web Companion. 2017 Presented at: WWW'17 Companion; April 03-07, 2017; Perth, Australia p. 1561-1566.
34. Yom-Tov E. Crowdsourced Health: How What You Do on the Internet Will Improve Medicine. Cambridge, Massachusetts: MIT Press; 2016.
35. Fond G, Gaman A, Brunel L, Haffen E, Llorca PM. Google Trends: Ready for real-time suicide prevention or just a Zeta-Jones effect? An exploratory study. Psychiatry Res 2015 Aug 30;228(3):913-917. [doi: 10.1016/j.psychres.2015.04.022] [Medline: 26003510]
36. Kristoufek L, Moat H, Preis T. Estimating suicide occurrence statistics using Google Trends. EPJ Data Sci 2016 Nov 8;5(1):32. [doi: 10.1140/epjds/s13688-016-0094-0]
37. National Crimes Record Bureau of India. 2014. Accidental Deaths & Suicides in India URL: http://ncrb.gov.in/StatPublications/ADSI/ADSI2014/adsi-2014%20full%20report.pdf [accessed 2018-02-20] [WebCite Cache ID 6xNMCUc9H]
38. Census 2011 India. 2011. States Census 2011 URL: http://www.census2011.co.in/states.php [accessed 2018-08-29] [WebCite Cache ID 722OwWT8h]
39. Press Information Bureau, Government of India. 2015 Jul. Per Capita National Income URL: http://pib.nic.in/newsite/PrintRelease.aspx?relid=123563 [accessed 2018-02-20] [WebCite Cache ID 6xNMFGLbF]
40. Ramani S. The Hindu. 2016 Aug 24. The India wide web URL: https://www.thehindu.com/sci-tech/technology/internet/The-India-wide-web/article14588938.ece [accessed 2018-09-21] [WebCite Cache ID 72au0i0L5]
41. Ministry of Statistics & Programme Implementation, Government of India. 2016. Statistical Year Book India URL: http://www.mospi.gov.in/statistical-year-book-india/2016/198 [accessed 2018-02-20] [WebCite Cache ID 6xNMM6PGq]
42. Sudhir T. Newslaundry. 2017 Oct 21. Mounting student suicides shock Telugu states URL: https://www.newslaundry.com/2017/10/21/student-suicides-entrance-exams-stress-telangana-andhra [accessed 2018-02-20] [WebCite Cache ID 6xNNpd5NN]
43. Chandra S. Linking Digital India As A Tool For Curbing Farmer Suicides – A Case Study Of Telangana State. SIJMD 2016 Apr 12;3(3):74-88. [doi: 10.19085/journal.sijmd030302]
44. Larsen ME, Nicholas J, Christensen H. A Systematic Assessment of Smartphone Tools for Suicide Prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]

Abbreviations

NCRB: National Crime Records Bureau
WHO: World Health Organization

Edited by G Eysenbach; submitted 20.02.18; peer-reviewed by Q Cheng, P Batterham; comments to author 02.08.18; revised version received 12.09.18; accepted 24.09.18; published 04.01.19.

Please cite as:
Adler N, Cattuto C, Kalimeri K, Paolotti D, Tizzoni M, Verhulst S, Yom-Tov E, Young A
How Search Engine Data Enhance the Understanding of Determinants of Suicide in India and Inform Prevention: Observational Study
J Med Internet Res 2019;21(1):e10179
URL: https://www.jmir.org/2019/1/e10179/
doi:10.2196/10179
PMID:30609976

Original Paper

Association Between Cancer Incidence and Mortality in Web-Based Data in China: Infodemiology Study

Chenjie Xu[1*], BAdmin; Yi Wang[2*], MSc; Hongxi Yang[1*], BAdmin; Jie Hou[3], MSc; Li Sun[4], MD; Xinyu Zhang[1], MD, PhD; Xinxi Cao[1], BAdmin; Yabing Hou[1], BAdmin; Lan Wang[4], PhD; Qiliang Cai[5], MD, PhD; Yaogang Wang[1], MD, PhD

[1]School of Public Health, Tianjin Medical University, Tianjin, China

[2]Tandon School of Engineering, New York University, New York, NY, United States

[3]School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China

[4]School of Nursing, Tianjin Medical University, Tianjin, China

[5]The Second Hospital of Tianjin Medical University, Tianjin Medical University, Tianjin, China

[*]these authors contributed equally

Corresponding Author:
Yaogang Wang, MD, PhD
School of Public Health
Tianjin Medical University
No 22, Qixiangtai Road, Heping District
Tianjin,
China
Phone: 86 13820046130
Email: wyg@tmu.edu.cn

Abstract

Background: Cancer poses a serious threat to the health of Chinese people, resulting in a major challenge for public health work. Today, people can obtain relevant information from not only medical workers in hospitals, but also the internet in any place in real-time. Search behaviors can reflect a population's awareness of cancer from a completely new perspective, which could be driven by the underlying cancer epidemiology. However, such Web-retrieved data are not yet well validated or understood.

Objective: This study aimed to explore whether a correlation exists between the incidence and mortality of cancers and normalized internet search volumes on the big data platform, Baidu. We also assessed whether the distribution of people who searched for specific types of cancer differed by gender. Finally, we determined whether there were regional disparities among people who searched the Web for cancer-related information.

Methods: Standard Boolean operators were used to choose search terms for each type of cancer. Spearman's correlation analysis was used to explore correlations among monthly search index values for each cancer type and their monthly incidence and mortality rates. We conducted cointegration analysis between search index data and incidence rates to examine whether a stable equilibrium existed between them. We also conducted cointegration analysis between search index data and mortality data.

Results: The monthly Baidu index was significantly correlated with cancer incidence rates for 26 of 28 cancers in China (lung cancer: r=.80, P<.001; liver cancer: r=.28, P=.016; stomach cancer: r=.50, P<.001; esophageal cancer: r=.50, P<.001; colorectal cancer: r=.81, P<.001; pancreatic cancer: r=.86, P<.001; breast cancer: r=.56, P<.001; brain and nervous system cancer: r=.63, P<.001; leukemia: r=.75, P<.001; Non-Hodgkin lymphoma: r=.88, P<.001; Hodgkin lymphoma: r=.91, P<.001; cervical cancer: r=.64, P<.001; prostate cancer: r=.67, P<.001; bladder cancer: r=.62, P<.001; gallbladder and biliary tract cancer: r=.88, P<.001; lip and oral cavity cancer: r=.88, P<.001; ovarian cancer: r=.58, P<.001; larynx cancer: r=.82, P<.001; kidney cancer: r=.73, P<.001; squamous cell carcinoma: r=.94, P<.001; multiple myeloma: r=.84, P<.001; thyroid cancer: r=.77, P<.001; malignant skin melanoma: r=.55, P<.001; mesothelioma: r=.79, P<.001; testicular cancer: r=.57, P<.001; basal cell carcinoma: r=.83, P<.001). The monthly Baidu index was significantly correlated with cancer mortality rates for 24 of 27 cancers. In terms of the whole population, the number of women who searched for cancer-related information has slowly risen over time. People aged 30-39 years were most likely to use search engines to retrieve cancer-related knowledge. East China had the highest Web search volumes for cancer.

Conclusions: Search behaviors indeed reflect public awareness of cancer from a different angle. Research on internet search behaviors could present an innovative and timely way to monitor and estimate cancer incidence and mortality rates, especially for cancers not included in national registries.

(J Med Internet Res 2019;21(1):e10677) doi:10.2196/10677

KEYWORDS

cancer; incidence; mortality; web-based data; internet searching

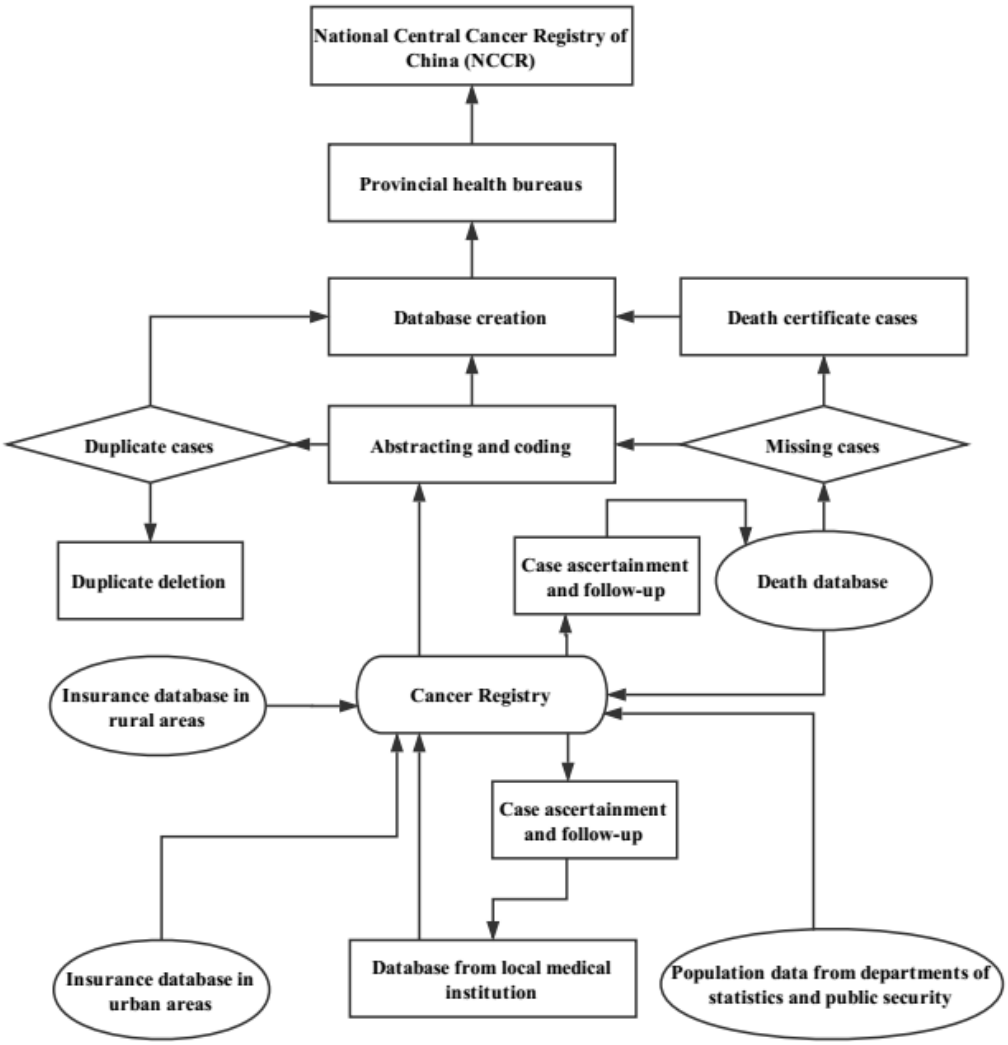
Introduction

Cancer affects people of all socioeconomic levels all over the world [1,2]. The global burden of cancer is increasing [3]. The population of China accounts for 19.3% of the global population, and the incidence of cancer accounts for 22% of global cancer incidence, ranking first in the world. Cancer deaths in China account for about 27% of global cancer deaths. Cancer mortality in China is also higher than the global average of 17% [4,5]. There is a growing demand for knowledge about cancers, but the registration of cancer cases in China requires complicated procedures (Figure 1) [6]. Traditional epidemiologic methods usually have a 3-year delay until incidence and mortality data are publicly reported due to the time required for data collection, compilation, quality control, and dissemination [7-9]. However, because nonmelanoma skin cancers (basal cell carcinoma and squamous cell carcinoma) are relatively nonlethal and curable by surgery, they are not covered by national surveillance and lack corresponding epidemiological data [10]. Given the inadequacy of traditional methods and the absence of data sources, internet search data can be used to estimate the characteristics of diseases [11].

Today, social media and medical forums are rapidly spreading, and internet users are increasingly exchanging health-related information. When people feel ill or have early symptoms, they may tend to first look for relevant health information on the internet for self-assessment. Some studies have improved the surveillance of epidemics and examined public interest in multiple health topics by monitoring the search behaviors of millions of users and conducting data mining through Google [12]. Other studies have tried to identify medication concerns, examine patient experience sentiments, and understand public perceptions by text mining social network data (eg, Facebook and Twitter) [13,14]. Studies have found that about 63% of cancer patients use the internet to retrieve cancer-related information [15-17]. A significant number of cancer patients utilize the internet to collect information about their respective diagnoses. A substantial number of cancer patients utilize the internet to gather information about their course of disease development [18]. It is becoming increasingly clear that the internet is a frequent source of information in our society for patients with cancer [19]. Compared with the past, at present, more Chinese internet users choose to retrieve information from the internet to obtain diagnosis and treatment information [20]. In addition to the patients themselves, friends and family members look for disease information using search engines, apps, and other resources, which are often designed to provide potentially helpful suggestions [21]. Studies in the United States have found a positive correlation between Google search volume and cancer incidence and mortality [22,23]. With the advancement of methodologies using “Big Data,” researchers are able to track diseases by the use of common internet search engines as a real-time tool [24]. Web search content is publicly available worldwide, providing valuable data for research, including health-related topics [25].

In this study, we tracked and monitored the Baidu index [26] and the search behaviors of Chinese internet users to explore public search interest in cancers. We also explored whether gender, age, and regional differences existed in search behaviors. We hypothesized that internet search volumes can reflect the disease characteristics of cancer (such as incidence and mortality) and provide an additional means of cancer surveillance in China.

Figure 1. Flow diagram of the cancer registration system.



Methods

Cancer Data

National-level incidence and mortality rates of cancers in China were obtained for the period 2011-2016 from the Global Burden of Disease database, which is publicly available [27]. For this study, we selected 28 types of cancer, including lung cancer, liver cancer, stomach cancer, esophageal cancer, colon and rectal cancer, pancreatic cancer, breast cancer, brain and nervous system cancer, cervical cancer, prostate cancer, nasopharynx cancer, bladder cancer, gallbladder and biliary tract cancer, lip and oral cavity cancer, ovarian cancer, larynx cancer, kidney cancer, testicular cancer, uterine cancer, thyroid cancer, multiple myeloma, leukemia, Non-Hodgkin lymphoma, malignant skin melanoma, Hodgkin lymphoma, mesothelioma, basal cell carcinoma, and squamous cell carcinoma. We also obtained the incidence and mortality rates of cancers according to gender, although mortality rates for basal cell carcinoma were missing. Due to the lack of monthly data for incidence and mortality rates, we used annual incidence and mortality rates for each cancer instead.

Web Search Data

This study mainly considered cancer search index values from Chinese search engines. The Baidu index was used as the entry point to launch the corresponding research [26]. Among Chinese search engine users (searches are usually conducted in Chinese), Baidu accounts for 92.1% of searches, followed by Haosou [28] and Sougou [29]. Regarding mobile search engines, the brand performance is the same: Baidu ranks first with 93.1% of the use rate. Baidu is a well-known Chinese search engine with powerful real-time functions [30]; it holds a strong position in China. Baidu is a very large information resource-sharing platform that Chinese netizens depend on. The Baidu index has proven to be a useful indicator of public interest in and awareness of health-related topics [31,32]. We assumed the Baidu index could best represent the retrieval preferences of Chinese internet users.

The Baidu index derives from search frequencies on the Baidu search engine; it is calculated and displayed based on the search volumes of specific keywords entered by users [33]. We entered the search terms for cancers according to the settings in the Baidu index to obtain the monthly total search index values of all cancers from January 2011 to December 2016. The daily

Baidu index is the weighted sum of the search frequency for a keyword based on its daily search volume on Baidu. The monthly Baidu search index value is the average of the total daily search index values in a month. We also obtained gender, age, and regional distribution data for people who retrieved cancer information online.

Search Terms

In this study, cancer awareness was examined on the basis of the general population's ability to seek information on or pay attention to the disease. Because Baidu is a Chinese search engine, the search terms are all expressed in Chinese characters. Given the diverse meanings of Chinese characters, in addition to their formal Chinese names, some cancers have various synonyms. All their formal Chinese names were referenced to the International Classification of Diseases for Oncology. Therefore, we selected both the formal Chinese names and common terms for various cancers while searching. Standard Boolean operators were used to combine terms. The search index value for each cancer could be incorporated into five keywords, and the selected terms were not searched in quotes. For most cancers, we used two or more search terms in Chinese to cover as many synonyms as possible.

The Baidu index covers the function of keyword analysis, which is the process of scientifically determining keywords based on the mode through which the searchers initiate a search request. According to the time period of the research (January 2011 to December 2016), the Baidu index system automatically analyzed the flow and trend of keywords imported in the Baidu search engine. We first entered the formal Chinese names of various cancers as keywords. The keyword analysis function automatically generated a corresponding number of related words and the search demand of the related words themselves. These words could be used as search terms to reflect people's retrieval needs. The function of keyword analysis helped us screen the search terms preliminary. We also conducted different retrieval methods for keyword selection to make the process more rigorous. We conducted comparative retrieval, cumulative retrieval, and combined retrieval for keywords and related words. Comparative retrieval aims to separate different keywords with commas among multiple words, which can realize the comparative query of keyword data. Cumulative retrieval indicates that among different keywords, different keywords are connected by a plus sign, and the addition of different keyword data can be realized. The aggregated data are presented as a combination of keywords. Combined retrieval is a combination of "comparative retrieval" and "cumulative retrieval." Subsequently, the search terms of each cancer can be determined.

For non-Hodgkin lymphoma, we also added the more common term "lymphoma" because that search term is twice as common as "non-Hodgkin lymphoma," and approximately 90% of lymphomas are non-Hodgkin lymphoma [34,35]. For prostate cancer, larynx cancer, Hodgkin lymphoma, and mesothelioma, we only used one search term, since their synonyms were not included in the Baidu index. We were unable to include their synonyms because they lacked unifying search terms with adequate search data for the analysis. Multimedia Appendix 1 shows all the search terms.

Statistical Analysis

First, we performed the Spearman correlation analysis to evaluate the relationship between the known cancer incidence and mortality rates for all cancer types and the Baidu index for the period 2011-2016. The distribution of the original variables is not required in the Spearman correlation analysis, as it is a nonparametric statistical method, and the scope of application is wider; thus, statistical significance was set as .05 (two-sided test).

Second, we used the Engel-Grange test to determine whether there was cointegration or long-term association between the three indicators. We defined the search index values and the incidence and mortality rates for each type of cancer over the past 6 years as time-series data. To eliminate heteroscedasticity in the time series, in the first step, we obtained the log version of the Baidu index, and incidence and mortality rates [36]. The advantage of this step is that data with large spacing can be converted into data with small spacing. Thereafter, we used unit root tests to examine whether the time series of the Baidu index for cancer searches and the time series for cancer incidence and mortality rates were stationary. If the three time series were all stationary at the same level, we estimated cointegration using ordinary least squares. We performed two types of cointegration analysis. Baidu index for cancer searches was used as the independent variable. The cancer incidence rate and cancer mortality rate were used as the dependent variable separately. In the third step, we used unit root test to test whether the residual series of the cointegrating regression model was stationary, which would show that the time series variables were cointegrated.

Statistical analysis was conducted using IBM SPSS (version 22.0, IBM Corporation, Armonk, NY), EViews (version 8, IHS Global Inc, London, United Kingdom), and R project (version 3.4, R Development Core Team, Vienna, Austria). We used Tableau (version 2018.3, Tableau Software, Seattle, WA) to conduct statistical analysis and create figures.

Results

Incidence and Mortality Rates of Cancers in China

We obtained the cancer incidence and mortality rates from 2011 to 2016 using data from the Global Burden of Disease database [27]. We illustrated the differences in the incidence and mortality rates of 28 cancers in China as well as the differences between men and women in 2016 (Figures 2, 3, and 4).

Figure 2. Incidence and mortality rates of cancers in China, 2016. Data were obtained from the Global Burden of Disease database. The blue signs represent incidence rates and the red signs represent mortality rates.

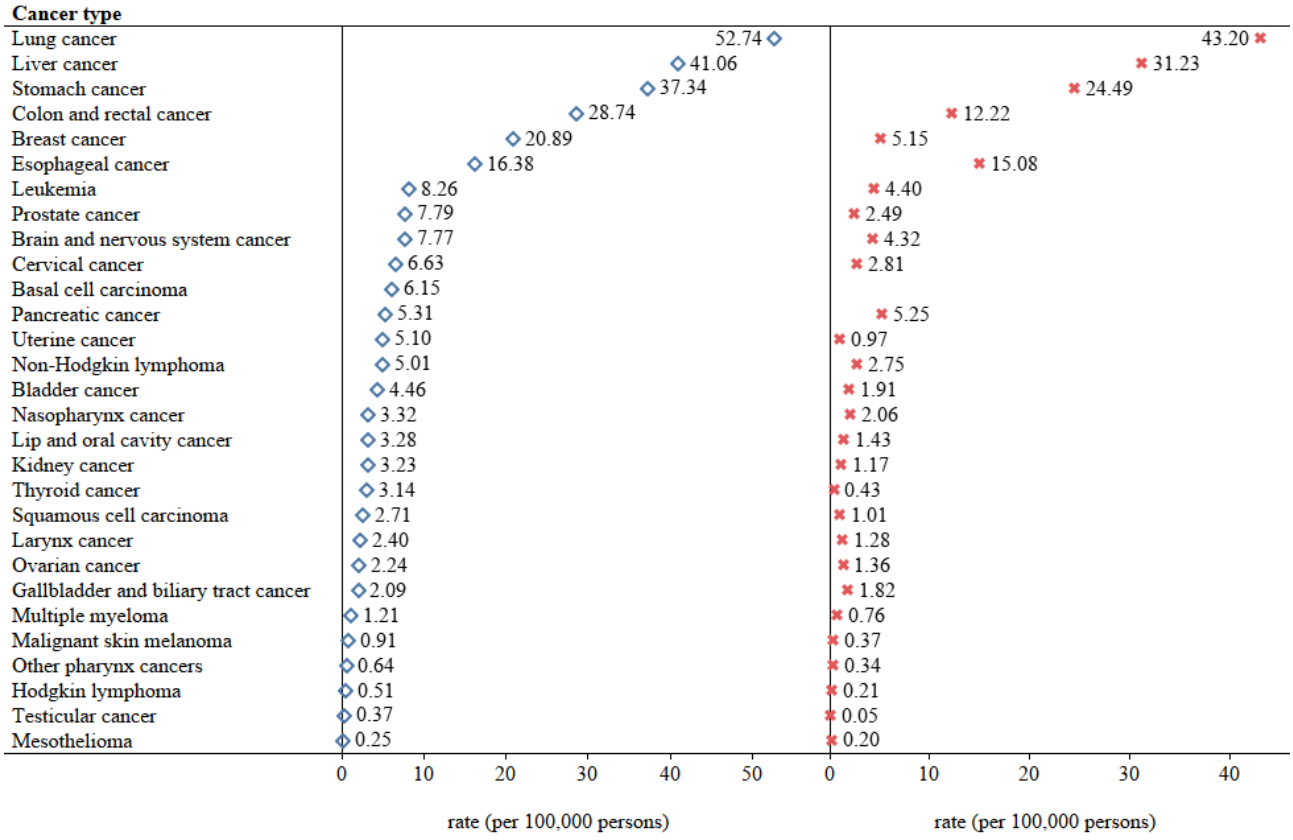


Figure 3. Incidence and mortality rates of cancers in China among men in 2016. The blue signs represent incidence rates and the red signs represent mortality rates.

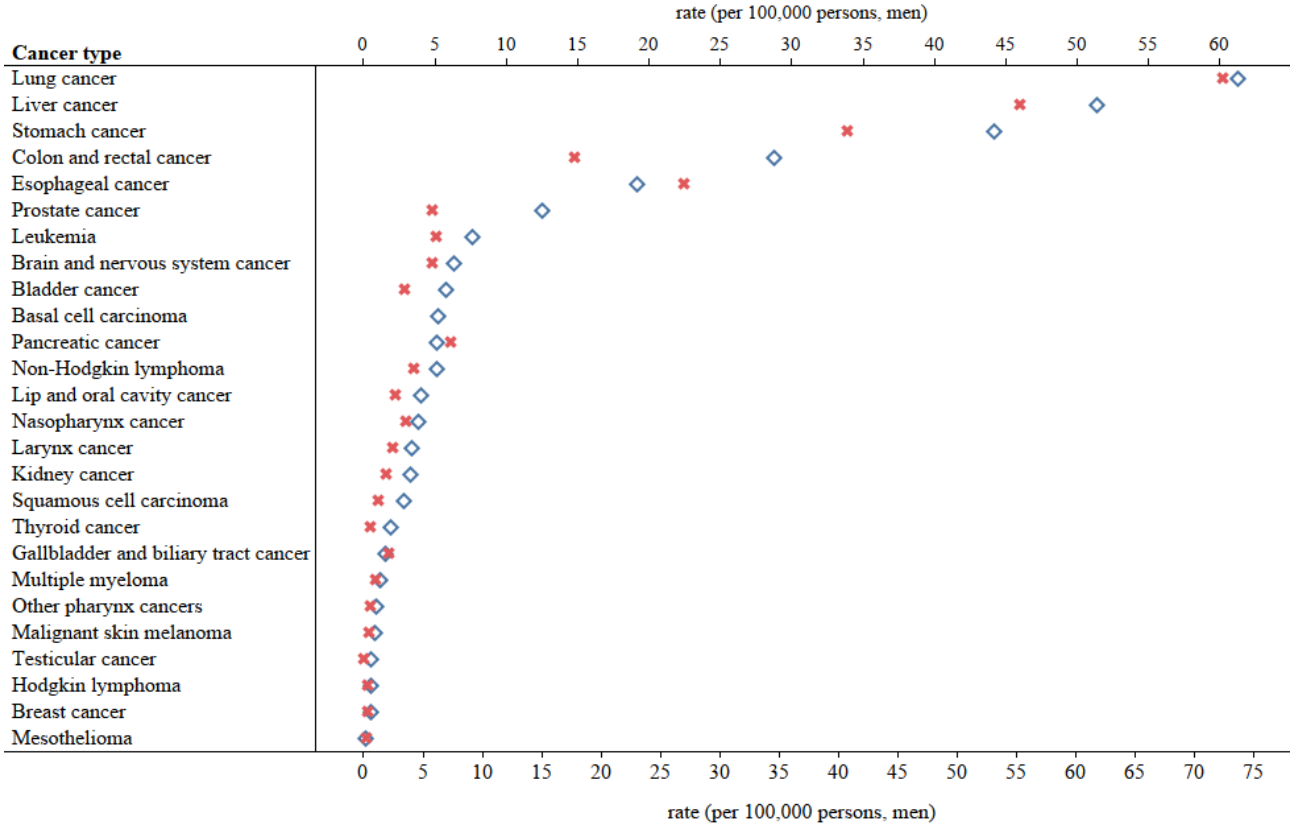
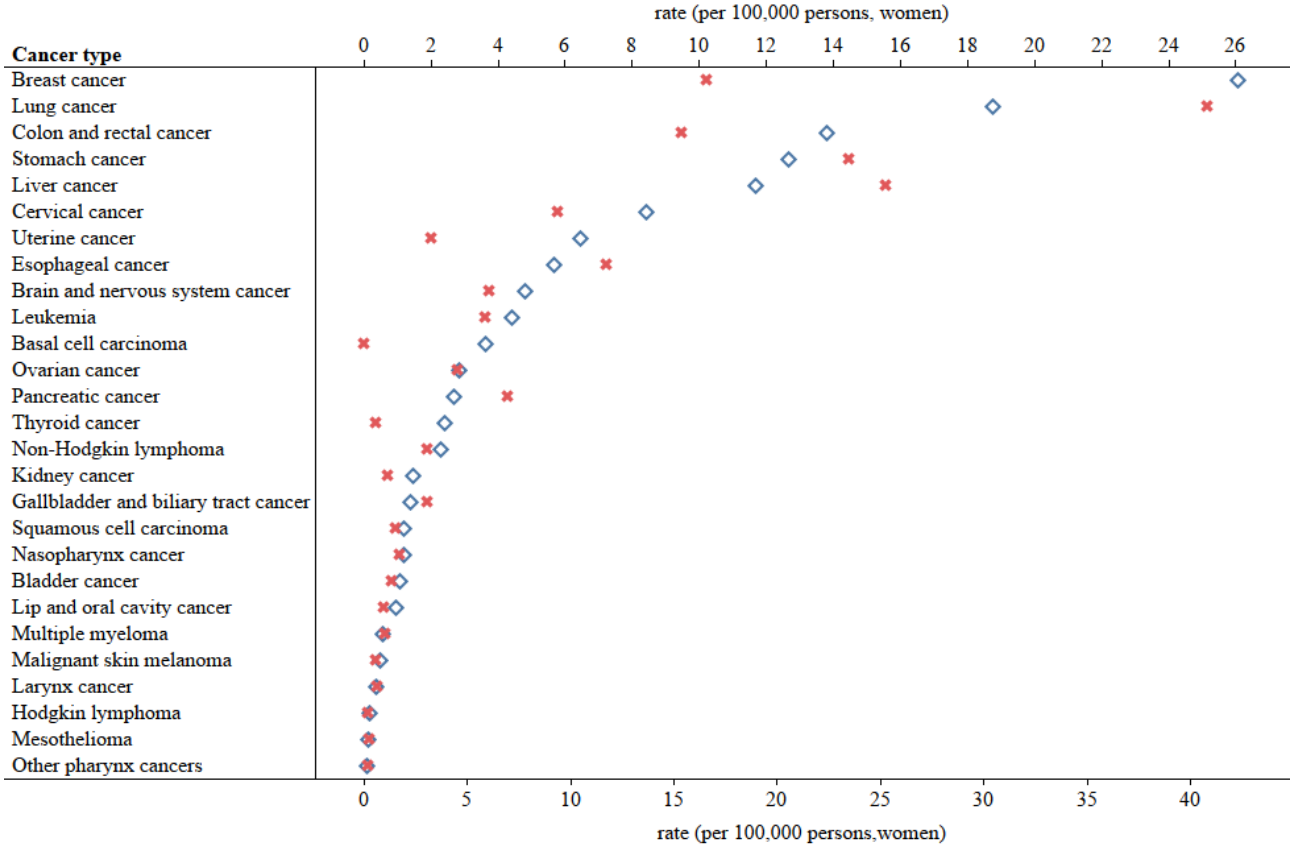


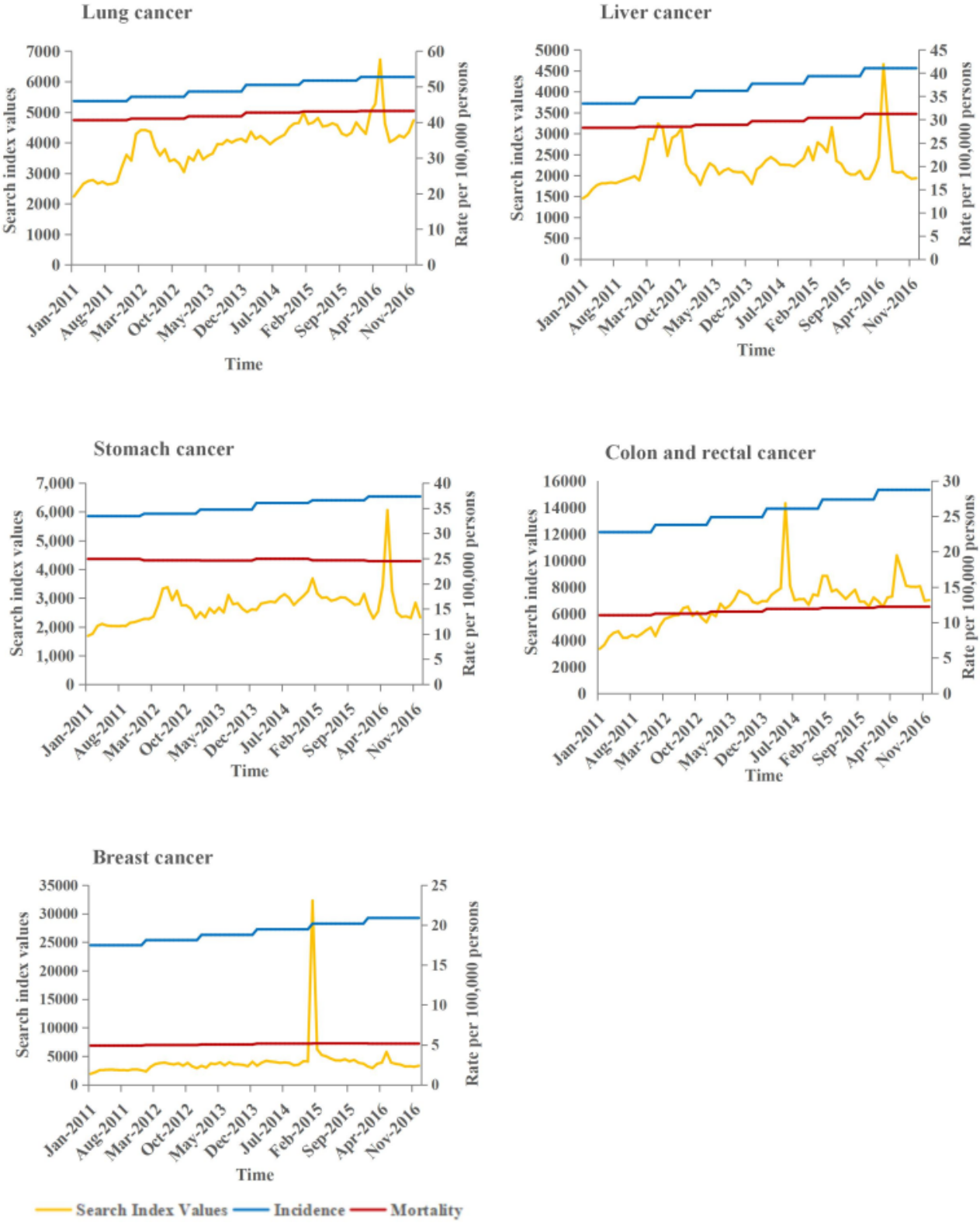
Figure 4. Incidence and mortality rates of cancers in China among women in 2016. The blue signs represent incidence rates and the red signs represent mortality rates.



Trends in Web-Based Data, Cancer Incidence, and Mortality Rates

Figure 5 shows a times series of the Baidu index and the incidence and mortality rates for the top five most common cancers in China. The remaining cancer types are shown in the Multimedia Appendix 2. The search index for these cancers is relatively flat at first, eventually showing a fluctuating trend over time. For multiple myeloma, part of the figure shows a "W" shape during this time, indicating significant fluctuation in the search index values. For mesothelioma, part of the figure shows a "V" shape, representing one search valley. For malignant skin melanoma, the monthly Baidu search index values showed a downward trend at first. From January 2015 to April 2016, the search trends of non-Hodgkin lymphoma and Hodgkin lymphoma showed consistent changes in volatility. Overall, searches for cancer terms showed an upward trend. At the same time, single or multiple peaks emerged in the fluctuation. The Baidu index data for breast cancer reached a peak in 2015; the search index value reached a maximum of 32,284 and the search frequency suddenly increased dramatically. A similar trend was observed for testicular cancer in November 2016, although the average search index was high. The trend values for uncommon cancers were relatively volatile.

Figure 5. Time series of search index values and incidence and mortality rates (top five most commonly occurring cancers).

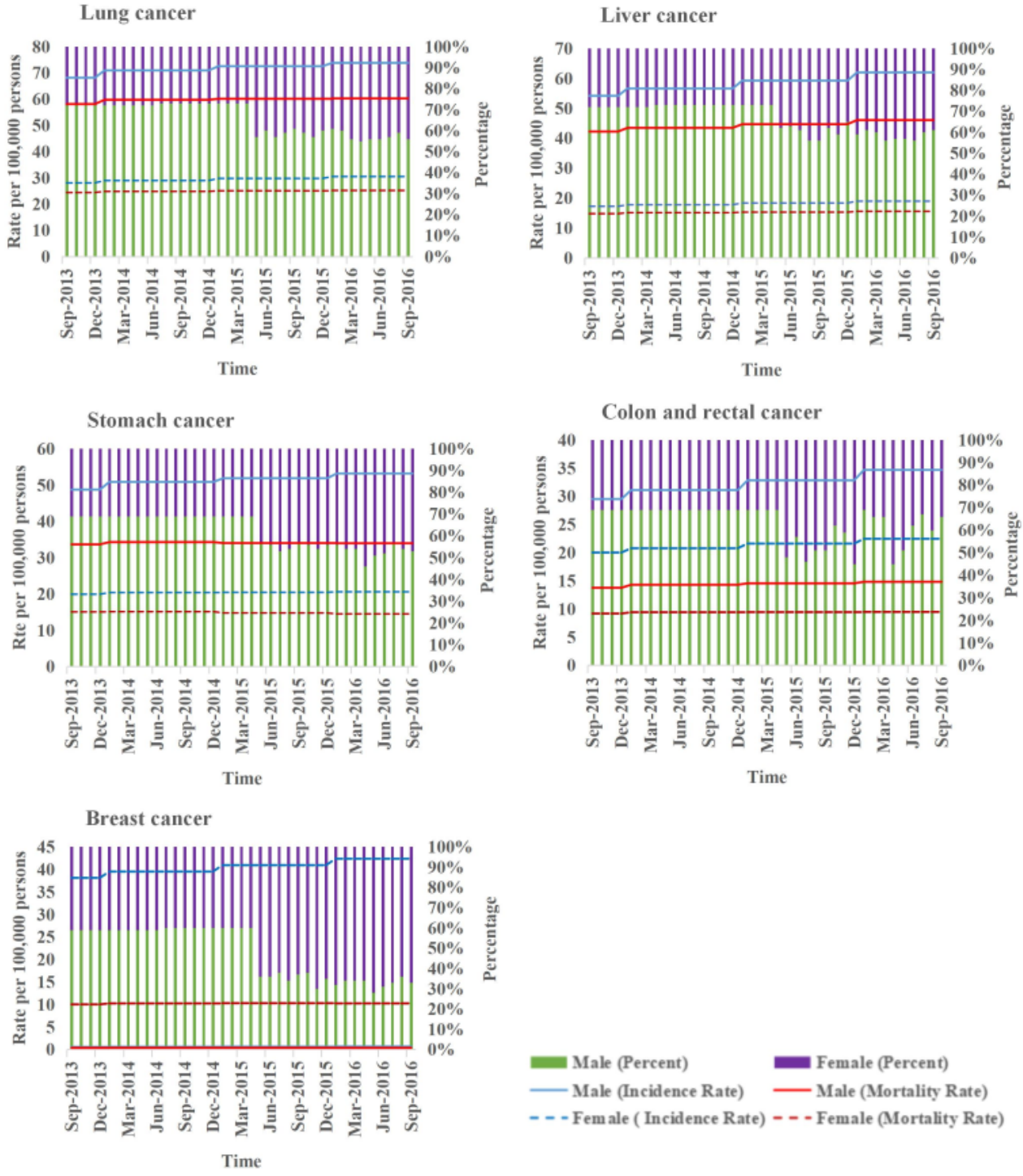


Gender Differences

For Hodgkin lymphoma, gender percentage data were missing before May 2015. The incidence and mortality rates of brain, nervous system, thyroid, gallbladder, and biliary tract cancers were found to be higher for women than for men. Incidence and mortality rates of female-specific cancers (breast, cervical, and uterine cancers) were also higher than those of male-specific cancers (prostate and testicular). Other cancers had higher incidence and mortality rates among men than among women. Incidence and mortality rates have increased annually among both men and female for lung cancer, liver cancer, colorectal cancer, pancreatic cancer, brain and nervous system cancer, non-Hodgkin lymphoma, prostate cancer, bladder cancer,

gallbladder and biliary tract cancer, lip and oral cavity cancer, ovarian cancer, kidney cancer, multiple myeloma, and malignant skin melanoma. Relatively, men paid more attention to search terms related to these cancers than women. In terms of the whole population, the number of women who searched for cancer-related information has slowly risen since 2015, while the number of men has shown a downward trend. This trend is even more obvious for female-specific cancers such as breast, cervical, ovarian, and uterine cancers. Initially, more men searched for terms related to breast cancer, but over time, an increasing number of women searched for such terms. More men paid attention to prostate and testicular cancers than women (Figure 6, Multimedia Appendix 2).

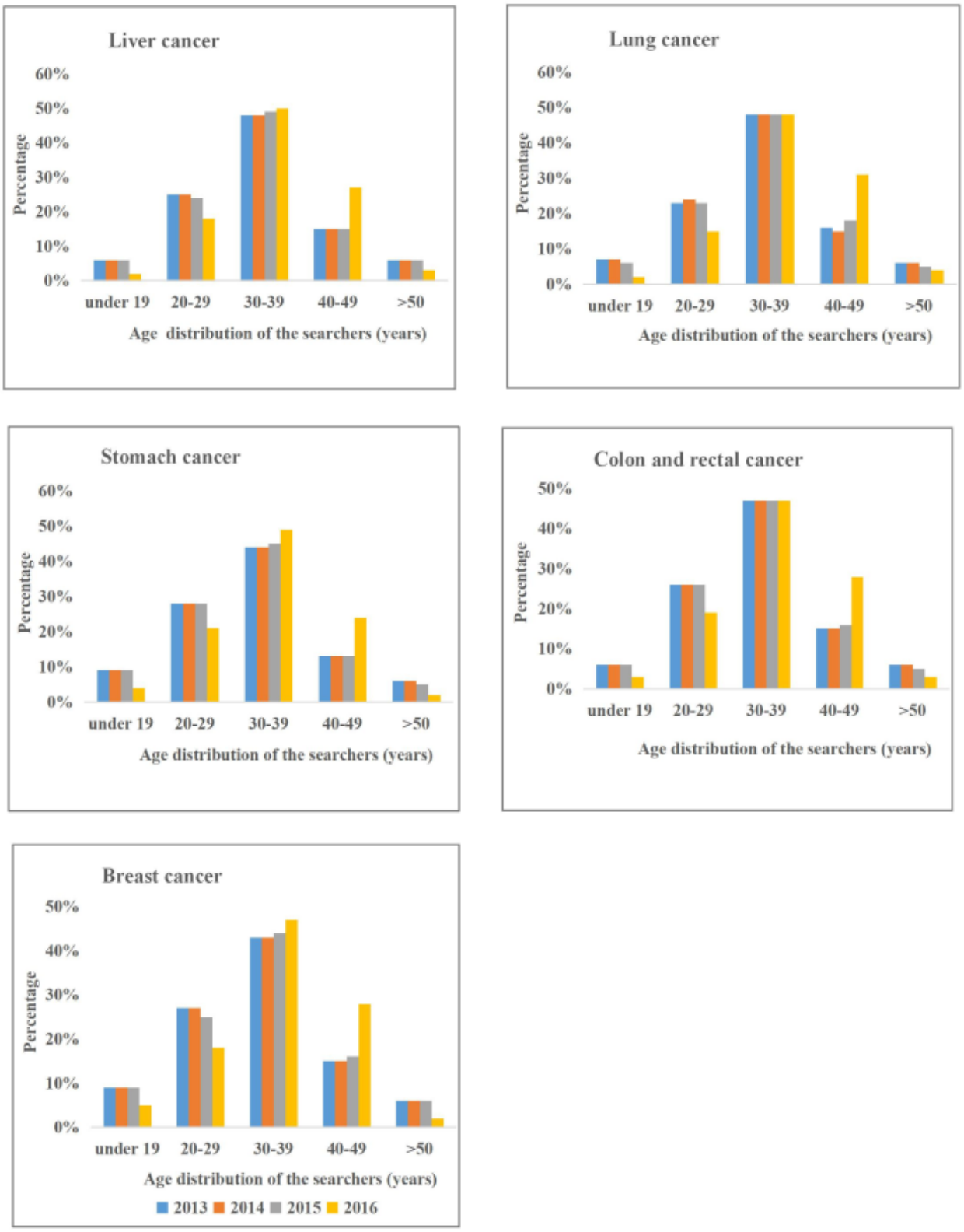
Figure 6. Incidence, mortality, and search distribution of cancers divided by gender (top five most commonly occurring cancers). The percentile chart represents the change from September 2013 to September 2016.



Age Distribution

Figure 7 and Multimedia Appendix 2 show the age distribution for cancer-related searches from 2013 to 2016 in China. As Figure 7 shows, the age group of 30-39 years was the largest search group for each cancer type, and the group aged over 50 years was the smallest. The proportion of the age groups of 30-39 years and 40-49 years increased over the last 4 years of the study period. The remaining age groups showed an opposite trend.

Figure 7. Age distribution of the searchers from 2013 to 2016 (top five most commonly occurring cancers).



Regional Distribution

Figure 8 shows the rankings of regional cancer search index values from 2013 to 2016 in 31 Chinese provinces and cities. Ranking was determined by the size of the web search volumes. In the heat maps, Guangdong Province shows the highest search index value and Tibet shows the lowest. For the top five cancer types, search values were the highest in eastern China (Shanghai, Jiangsu Province, Zhejiang Province, Anhui Province, Fujian Province, Jiangxi Province, and Shandong Province) and the lowest in northwestern China (Shaanxi Province, Gansu Province, Qinghai Province, the Ningxia Hui autonomous region, and the Xinjiang Uygur autonomous region). North, south, central, southwest, and northeast China were ranked second to sixth, respectively, in the search index values.

Figure 8. Ranking of regional distribution of the online searchers from 2013 to 2016 (top five most commonly occurring cancers) in mainland China. AH: Anhui, BJ: Beijing, FJ: Fujian, GS: Gansu, GD: Guangdong, GX: Guangxi, GZ: Guizhou, HI: Hainan, HE: Hebei, HA: Henan, HL: Heilongjiang, HB: Hubei, HN: Hunan, JL: Jilin, JS: Jiangsu, JX: Jiangxi, LN: Liaoning, NM: Inner Mongoria, NX: Ningxia, QH: Qinghai, SD: Shandong, SX: Shanxi, SN: Shaanxi, SH: Shanghai, SC: Sichuan, TJ: Tianjing, XZ: Tibet, XJ: Xinjiang, YN: Yunnan, ZJ: Zhejiang, CQ: Chongqing.

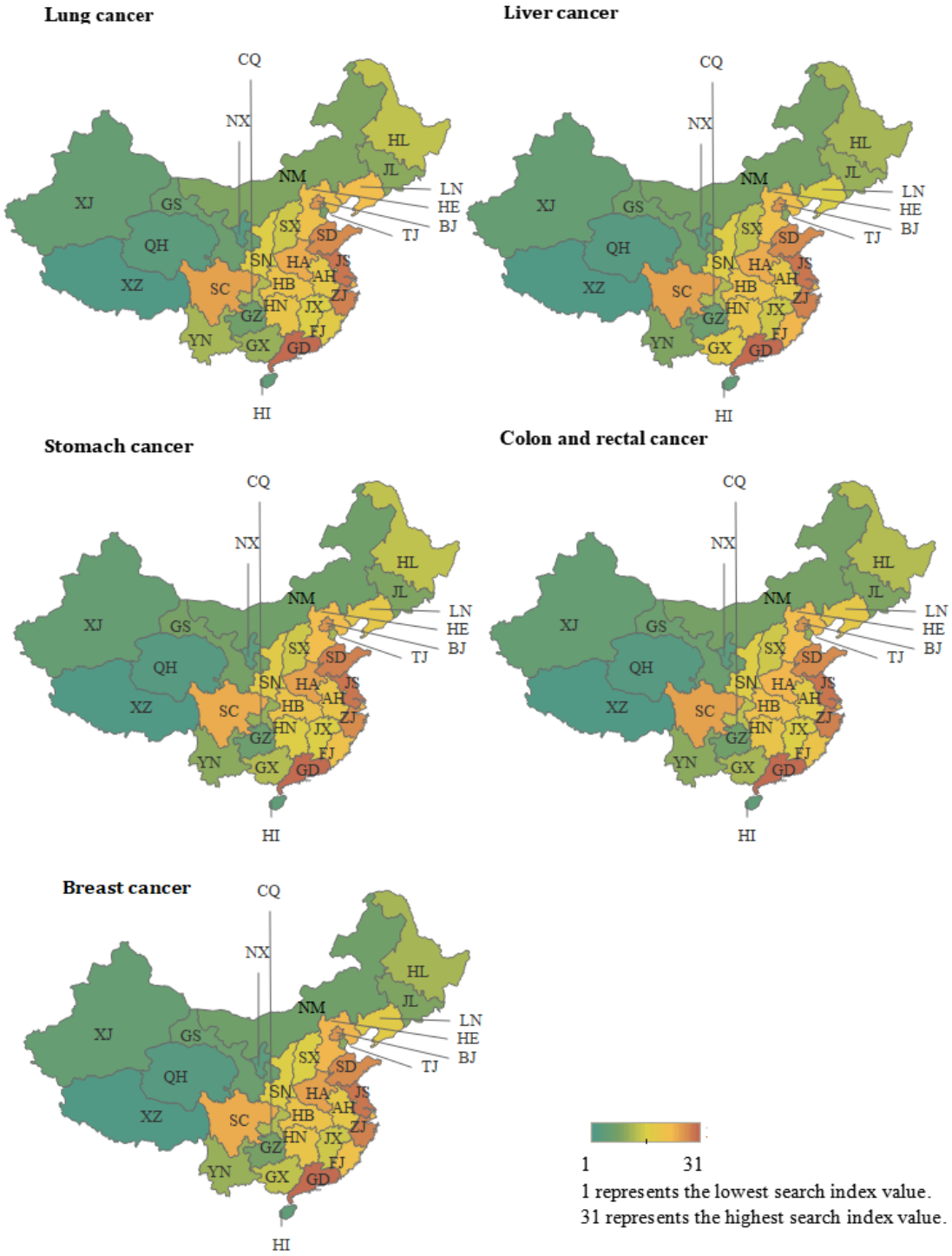


Table 1. Correlation coefficients between search index values, incidence rate of cancers, and mortality rate of cancers.

Cancer	Correlation between search index values and incidence rate		Correlation between search index values and mortality rate	
	r (correlation coefficient)	*P* value	*r* (correlation coefficient)	*P* value
Lung cancer	0.80	<.001	0.80	<.001
Liver cancer	0.28	.02	0.28	.02
Stomach cancer	0.50	<.001	0.02	.88
Esophageal cancer	0.50	<.001	0.21	.08
Colon and rectal cancer	0.81	<.001	0.81	<.001
Pancreatic cancer	0.86	<.001	0.86	<.001
Breast cancer	0.56	<.001	0.76	<.001
Leukemia	0.75	<.001	-0.70	<.001
Brain and nervous system cancer	0.63	<.001	0.63	<.001
Cervical cancer	0.64	<.001	0.65	<.001
Non-Hodgkin lymphoma	0.88	<.001	0.88	<.001
Prostate cancer	0.67	<.001	0.67	<.001
Nasopharynx cancer	0.08	.51	0.44	<.001
Bladder cancer	0.62	<.001	0.62	<.001
Gallbladder and biliary tract cancer	0.88	<.001	0.88	<.001
Lip and oral cavity cancer	0.88	<.001	0.88	<.001
Ovarian cancer	0.58	<.001	0.58	<.001
Larynx cancer	0.82	<.001	0.74	<.001
Kidney cancer	0.73	<.001	0.73	<.001
Squamous cell carcinoma	0.94	<.001	0.87	<.001
Uterine cancer	0.04	.73	-0.42	<.001
Multiple myeloma	0.84	<.001	0.84	<.001
Thyroid cancer	0.77	<.001	0.77	<.001
Malignant skin melanoma	0.55	<.001	0.55	<.001
Hodgkin lymphoma	0.91	<.001	-0.91	<.001
Mesothelioma	0.79	<.001	0.79	<.001
Testicular cancer	0.57	<.001	-0.08	.48
Basal cell carcinoma	0.83	<.001	__[a]	__[a]

[a]Not available.

Correlation Analysis

Table 1 shows the correlation coefficients between actual incidence rates and the relative Baidu index for 28 cancers in China. We found statistically significant correlations between incidence rates and the relative Baidu index for 26 cancers (nasopharynx cancer and uterine cancer did not show such correlations).

Table 1 also shows the correlation coefficients between actual mortality rates and the relative Baidu index for these cancers. Stomach cancer, esophageal cancer, and testicular cancer did not show statistically significant correlations with mortality rates. For leukemia, uterine cancer, and Hodgkin lymphoma, the relative Baidu index was negatively correlated with cancer mortality rates.

Cointegration Analysis

Augmented Dickey-Fuller unit root test was used to examine the stationarity of the time series. Schwarz information criterion was used to determine lag length automatically. We first made logarithmic changes to the three indexes. After transformation, the series were all stationary at the first difference (Multimedia Appendix 3). Since the series were found to be stationary at the same level, the three variables satisfied the precondition of cointegration and were checked for a long-term cointegration relationship. The result of the cointegration (Engle-Granger) test showed cointegration between variables for the Baidu index

and incidence rates at the first difference (Multimedia Appendix 4). The cointegration test also showed cointegration between variables for the Baidu index and mortality rates at the first difference (Multimedia Appendix 5).

Discussion

Principal Findings

For most cancers, the Baidu index was positively correlated with cancer incidence rates. For several cancers including lung cancer, liver cancer, stomach cancer, colon and rectal cancer, breast cancer, prostate cancer, brain and nervous system cancer, cervical cancer, pancreatic cancer, non-Hodgkin lymphoma, bladder cancer, nasopharynx cancer, lip and oral cavity cancer, kidney cancer, thyroid cancer, squamous cell carcinoma, larynx cancer, ovarian cancer, gallbladder and biliary tract cancer, multiple myeloma, malignant skin melanoma and mesothelioma, the Baidu index was positively correlated with cancer mortality rates. The results suggest that the search engine data can reflect actual prevalence to some extent. Such data sources might be particularly useful when real-time information is required or missing (eg, mortality rate of basal cell carcinoma is lacking), considering that there is often a lag of several years in the publication of cancer registration data. The results of this study suggest that we should study and make use of Web-based data and publicly available information regarding people's interest in health topics to estimate cancer trends. Although most cancers examined in this study showed statistically significant correlations of the Baidu index with incidence and mortality rates, nasopharynx, uterine, stomach, esophageal, and testicular cancers did not show such correlations. This is probably attributable to various public health-related phenomena that may increase search volumes independent of disease metrics, such as the National Cancer Prevention Week held by the China Anti-Cancer Association (April of each year) or appearance of reports of cancer among public figures. After launch of a public health campaign for a disease, the information-search behavior associated with the disease will also increase [37]. For example, during the US annual breast cancer awareness campaign in October, online activity was stimulated and the number of Google searchers for "breast cancer" increased significantly [22,38]. For leukemia, uterine cancer, and Hodgkin lymphoma, the relative Baidu index was negatively correlated with cancer mortality rates. The possible reason for this might be the differences in the amount of data. The search index values for these three cancers gradually increased with time and showed large absolute values (Multimedia Appendix 2), and their mortality rates were low and stable over time. The trend of search engine search terms may be affected by other factors such as public panic [39]. Owing to the convenient use of the internet and the reports on internet media, people are more familiar with the three abovementioned cancers. Similarly, interest in breast cancer increased in January 2015 in China, perhaps because of the death of the well-known singer Beina Yao due to breast cancer.

In terms of the gender distribution of the search population, there were initially more men than women in our study. This could be attributable to gender differences in the disease burden pattern. For example, men are more susceptible than women to various deadly diseases, including cancer [40,41]. The gradual increase in the percentage of women searching for cancer-related information (especially for breast, ovarian, and uterine cancers) reflects increased health awareness among women. For other cancers, the gender structure of internet users tended to be balanced and basically consistent with the sex ratio of the population. In the process of obtaining search index data, we also found that people aged over 50 years were most prone to cancer among all age groups. The proportion of this age group among people who search online for cancer-related information is low, because they are less familiar with mobile devices and internet use [42]. Therefore, cancer-prevention initiatives should pay more attention to this age group to help them understand the relevant information. China is a vast and diverse country, with a population of more than 1.3 billion people. Regional differences were also found in search volumes, which could stem from regional disparities in demographic and socioeconomic conditions, education, and health literacy. For example, eastern China ranks first in search volumes, whereas less developed areas such as the northwest ranked last. People in densely populated and economically developed cities in eastern China had better internet access and higher health awareness. They search for health information more frequently than people in sparsely populated and developing cities. Local authorities should make efforts to ensure that online health information is accessible to the public, especially in economically underdeveloped areas.

Given the nonstandard treatments and other related issues, cancer diagnoses in China are generally made late, and the survival rates are not high [7]. Establishing effective cancer-control measures has thus become an important public health issue in China. Studies have shown that tracking and monitoring search index values as well as text mining on social media can provide new ways to study public concerns about cancer and information-seeking behaviors [12]. Web search content could provide valuable data for research on cancer-related topics [43,44].

Norman and Skinner defined eHealth literacy as "The ability to search, understand, and evaluate health information on electronic resources, and harnessing the information they receive to address and solve health problems". As the content of health literacy continued to expand, Norman and Skinner proposed that electronic health literacy is a combination of different abilities, which can be divided into two types: traditional literacy vs computing ability, media literacy, and information literacy. Computer literacy, scientific literacy, and health literacy refer to the ability to deal specifically with problems in specific areas [45,46]. At present, there is little research in China on cancer-related electronic health literacy (or health literacy, in general) despite the fact that public retrieval of cancer-related health information indirectly reflects individual levels of electronic health literacy. It is necessary to enhance the efficiency of prevention and early diagnosis for patients with cancer or the general population by online information transmission. Collection of real-time relevant search data from search engines provides a new way for cancer prevention and control.

Research on attention paid to health information in the Web as well as population characteristics can help estimate some indicators of diseases among the population, which can help improve the allocation of health resources and implementation of effective public health measures. This could also help medical providers who are facing various challenges including understanding characteristics of patients who use the internet, the reasons for utilizing the internet, and the effectiveness and security of websites currently providing health-related information to patients.

Strengths and Limitations

Previous studies have mainly used Google Trends [47] or the Baidu index to analyze the burden of epidemic diseases and predict their trends. This is the first study to explore the associations between online interest, cancer incidence, and mortality rates in China.

This study has some limitations. The use of Baidu search data to estimate disease metrics might not be completely generalizable, since the data are restricted to those with access to the internet. We were also unable to determine the types of internet users or which stakeholders were responsible for search activities. Given China's vast size and large population, the registry of cancer statistics is usually lagging and not comprehensive. We could not obtain timely data on the monthly incidence and mortality rates of all cancers in China. Use of search index values from a popular internet search engine can only account for a small portion of changes in incidence and mortality rates of cancers, which are also greatly affected by public health activities. Studying search engine data is inevitably restricted by these random factors; this is an unavoidable limitation in such research. We hope to find ways to identify and reduce bias in search engine data before we utilize Web-based data to provide useful information for cancer surveillance, evaluation of public cancer awareness, and education programs.

Conclusions

Owing to the widespread proliferation of internet technology, all kinds of people make use of the internet. In the medical field, it is often intended to prompt informed conversations with clinical professionals and suggest potentially helpful resources to patients or other people. Indeed, this study found a correlation between search index values and the incidence and mortality rates for most types of cancers. In a way, search behaviors and volumes can reflect the public awareness of cancer. Therefore, an advanced understanding of search behaviors could augment traditional epidemiologic surveillance and help achieve the goal of cancer prevention and control. It will be beneficial for us to pay attention to internet search data, especially when registry data are insufficient or lagging.

Acknowledgments

This research was funded by National Natural Science Foundation of China (#91746205; #71673199).

Authors' Contributions

YW developed the original research idea for the study and directed the study. All authors conducted the analysis and interpreted the results. CX, YW, and HY developed the first manuscript draft. All authors critically revised the manuscript. All authors critically reviewed and contributed to the final version and approved it. All authors had full access to the study data. YW had the final responsibility of the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms of 28 cancers in Chinese.

[PDF File (Adobe PDF File), 87KB - jmir_v21i1e10677_app1.pdf]

Multimedia Appendix 2

Relevant figures of the remaining types of cancers.

[PDF File (Adobe PDF File), 6MB - jmir_v21i1e10677_app2.pdf]

Multimedia Appendix 3

Results of unit root tests for the time series of monthly Baidu index, incidence rate, and mortality rate for each cancer type.

[PDF File (Adobe PDF File), 68KB - jmir_v21i1e10677_app3.pdf]

Multimedia Appendix 4

Results of the cointegration test of the two-time series of monthly Baidu indexes and incidence rates of cancers.

[PDF File (Adobe PDF File), 67KB - jmir_v21i1e10677_app4.pdf]

Multimedia Appendix 5

Results of the cointegration test of the two-time series of monthly Baidu indexes and mortality rates of cancers.

[PDF File (Adobe PDF File), 66KB - jmir_v21i1e10677_app5.pdf]

References

1. Meyrowitsch DW, Bygbjerg IC. Global burden of disease--a race against time. Dan Med Bull 2007 Feb;54(1):32-34. [Medline: 17349218]
2. World Health Organization. World health statistics 2017: Monitoring health for the SDGs. Geneva, Switzerland: WHO; 2017. URL: https://www.who.int/gho/publications/world_health_statistics/2017/en/ [accessed 2019-01-09] [WebCite Cache ID 75ITIFM5R]
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018 Jan;68(1):7-30 [FREE Full text] [doi: 10.3322/caac.21442] [Medline: 29313949]
4. Thun MJ, DeLancey JO, Center MM, Jemal A, Ward EM. The global burden of cancer: priorities for prevention. Carcinogenesis 2010 Jan;31(1):100-110 [FREE Full text] [doi: 10.1093/carcin/bgp263] [Medline: 19934210]
5. 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] [Medline: 25651787]
6. He J, Chen WQ. China cancer registry annual report 2017. Beijing, China: People's Medical Publishing House; 2018.
7. Chen WQ, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66(2):115-132 [FREE Full text] [doi: 10.3322/caac.21338] [Medline: 26808342]
8. Wei K, Chen W, Zhang S, Liang Z, Zheng R, Ou Z. Cancer registration in the Peoples Republic of China. Asian Pac J Cancer Prev 2012;13(8):4209-4214. [Medline: 23098534]
9. Wei K, Liang Z, Liu J, Wang X. [History of cancer registration in china]. Zhonghua Yi Shi Za Zhi 2012 Jan;42(1):21-25. [Medline: 22613477]
10. Liu-Smith F, Jia J, Zheng Y. UV-Induced Molecular Signaling Differences in Melanoma and Non-melanoma Skin Cancer. Adv Exp Med Biol 2017;996:27-40. [doi: 10.1007/978-3-319-56017-5_3] [Medline: 29124688]
11. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564 [FREE Full text] [doi: 10.1086/630200] [Medline: 19845471]
12. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
13. Polgreen PM, Chen Y, Pennock DM, Nelson FD. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 1;47(11):1443-1448 [FREE Full text] [doi: 10.1086/593098] [Medline: 18954267]
14. Park SH, Hong SH. Identification of Primary Medication Concerns Regarding Thyroid Hormone Replacement Therapy From Online Patient Medication Reviews: Text Mining of Social Network Data. J Med Internet Res 2018 Oct 24;20(10):e11085 [FREE Full text] [doi: 10.2196/11085] [Medline: 30355555]
15. Arora NK, Hesse BW, Rimer BK, Viswanath K, Clayman ML, Croyle RT. Frustrated and confused: the American public rates its cancer-related information-seeking experiences. J Gen Intern Med 2008 Mar;23(3):223-228 [FREE Full text] [doi: 10.1007/s11606-007-0406-y] [Medline: 17922166]
16. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the Internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [Medline: 11929503]
17. Kowalski C, Kahana E, Kuhr K, Ansmann L, Pfaff H. Changes over time in the utilization of disease-related Internet information in newly diagnosed breast cancer patients 2007 to 2013. J Med Internet Res 2014 Aug 26;16(8):e195 [FREE Full text] [doi: 10.2196/jmir.3289] [Medline: 25158744]
18. Castleton K, Fong T, Wang-Gillam A, Waqar MA, Jeffe DB, Kehlenbrink L, et al. A survey of Internet utilization among patients with cancer. Support Care Cancer 2011 Aug;19(8):1183-1190. [doi: 10.1007/s00520-010-0935-5] [Medline: 20556435]
19. Basch EM, Thaler HT, Shi W, Yakren S, Schrag D. Use of information resources by patients with cancer and their companions. Cancer 2004 Jun 01;100(11):2476-2483 [FREE Full text] [doi: 10.1002/cncr.20261] [Medline: 15160355]
20. Ji Z, Zhang Y, Xu J, Chen X, Wu Y, Xu H. Comparing Cancer Information Needs for Consumers in the US and China. Stud Health Technol Inform 2017;245:126-130 [FREE Full text] [Medline: 29295066]
21. Ayantunde AA, Welch NT, Parsons SL. A survey of patient satisfaction and use of the Internet for health information. Int J Clin Pract 2007 Mar;61(3):458-462. [doi: 10.1111/j.1742-1241.2006.01094.x] [Medline: 17313614]
22. Wehner MR, Nead KT, Linos E. Correlation Among Cancer Incidence and Mortality Rates and Internet Searches in the United States. JAMA Dermatol 2017 Dec 01;153(9):911-914 [FREE Full text] [doi: 10.1001/jamadermatol.2017.1870] [Medline: 28658470]

23. Phillips CA, Barz LA, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship Between State-Level Google Online Search Volume and Cancer Incidence in the United States: Retrospective Study. J Med Internet Res 2018 Jan 08;20(1):e6 [FREE Full text] [doi: 10.2196/jmir.8870] [Medline: 29311051]
24. Huang X, Baade P, Youlden DR, Youl PH, Hu W, Kimlin MG. Google as a cancer control tool in Queensland. BMC Cancer 2017 Dec 04;17(1):816 [FREE Full text] [doi: 10.1186/s12885-017-3828-x] [Medline: 29202718]
25. DeJohn AD, Schulz EE, Pearson AL, Lachmar EM, Wittenborn AK. Identifying and Understanding Communities Using Twitter to Connect About Depression: Cross-Sectional Study. JMIR Ment Health 2018 Nov 05;5(4):e61 [FREE Full text] [doi: 10.2196/mental.9533] [Medline: 30401662]
26. Baidu. URL: https://www.baidu.com/ [accessed 2018-04-04] [WebCite Cache ID 6yOv8CJvs]
27. GHDx. GBD Results Tool URL: http://ghdx.healthdata.org/gbd-results-tool [accessed 2018-04-04] [WebCite Cache ID 6yOsRvLfq]
28. 360 Search. URL: https://www.so.com/ [accessed 2019-01-23] [WebCite Cache ID 75e0rW7xD]
29. Sogou. URL: https://www.sogou.com/ [accessed 2019-01-23] [WebCite Cache ID 75e0zgeO0]
30. China Internet Network Information Center. Chinese Internet users search behavior study. Beijing, China; 2014. URL: http://www.cnnic.cn/hlwfzyj/hlwmrtj/201410/t20141017_49359.htm [accessed 2019-01-09] [WebCite Cache ID 75IWlqvRn]
31. Li Z, Liu T, Zhu G, Lin H, Zhang Y, He J, et al. Dengue Baidu Search Index data can improve the prediction of local dengue epidemic: A case study in Guangzhou, China. PLoS Negl Trop Dis 2017 Dec;11(3):e0005354 [FREE Full text] [doi: 10.1371/journal.pntd.0005354] [Medline: 28263988]
32. Yang H, Li S, Sun L, Zhang X, Hou J, Wang Y. Effects of the Ambient Fine Particulate Matter on Public Awareness of Lung Cancer Risk in China: Evidence from the Internet-Based Big Data Platform. JMIR Public Health Surveill 2017 Oct 03;3(4):e64 [FREE Full text] [doi: 10.2196/publichealth.8078] [Medline: 28974484]
33. Baidu. Baidu Index URL: https://index.baidu.com/ [accessed 2018-04-04] [WebCite Cache ID 6yOtOa7p9]
34. Tang F, Min L, Ye Y, Tang B, Zhou Y, Zhang W, et al. Classic Hodgkin lymphoma in pelvis: A case report highlights diagnosis and treatment challenges. Medicine (Baltimore) 2017 Sep;96(39):e8196 [FREE Full text] [doi: 10.1097/MD.0000000000008196] [Medline: 28953683]
35. Sabattini E, Bacci F, Sagramoso C, Pileri SA. WHO classification of tumours of haematopoietic and lymphoid tissues in 2008: an overview. Pathologica 2010 Jun;102(3):83-87. [Medline: 21171509]
36. Yi DH. Shuju fenxi yu EViews Yingyong. Beijing: China Renmin University Press; 2014.
37. Ling R, Lee J. Disease Monitoring and Health Campaign Evaluation Using Google Search Activities for HIV and AIDS, Stroke, Colorectal Cancer, and Marijuana Use in Canada: A Retrospective Observational Study. JMIR Public Health Surveill 2016;12(2):e156 [FREE Full text] [doi: 10.2196/publichealth.6504] [Medline: 27733330]
38. Glynn RW, Kelly JC, Coffey N, Sweeney KJ, Kerin MJ. The effect of breast cancer awareness month on internet search activity--a comparison with awareness campaigns for lung and prostate cancer. BMC Cancer 2011;11:442 [FREE Full text] [doi: 10.1186/1471-2407-11-442] [Medline: 21993136]
39. Zhou X, Shen H. Notifiable infectious disease surveillance with data collected by search engine. J Zhejiang Univ Sci C 2010;11(4):241-248. [doi: 10.1631/jzus.C0910371]
40. GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392(10159):1859-1922. [Medline: 30415748]
41. Courtenay WH. Constructions of masculinity and their influence on men's well-being: a theory of gender and health. Soc Sci Med 2000 May;50(10):401. [Medline: 10741575]
42. China Internet Network Information Center. CNNIC. Beijing; 2018. The 41th China Statistical Report on Internet Development URL: http://cnnic.cn/gywm/xwzx/rdxw/201801/t20180131_70188.htm [accessed 2018-04-03] [WebCite Cache ID 6yOupfhiY]
43. Khan AM, Khorana AA. Blood. 2016. Decrease in Awareness of Hematologic Malignancies in the United States: Temporal Analysis of Google Trends Search Data from 2004 to 2015 URL: http://www.bloodjournal.org/content/128/22/3565?sso-checked=true [accessed 2019-01-23] [WebCite Cache ID 75e5wLHHg]
44. Khan AM, Khorana AA. Estimating the outpatient burden of venous thromboembolism in the United States: an analysis of Google Trends data from 2004 to 2015. Blood 2015;126(23):4453 [FREE Full text]
45. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
46. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
47. Google Trends. URL: https://trends.google.com/trends/?geo=US [accessed 2019-01-23] [WebCite Cache ID 75e1rdRBY]

Edited by G Eysenbach; submitted 08.05.18; peer-reviewed by P Wark, W Lu, A Mavragani; comments to author 10.11.18; revised version received 04.12.18; accepted 06.01.19; published 29.01.19.

Please cite as:
Xu C, Wang Y, Yang H, Hou J, Sun L, Zhang X, Cao X, Hou Y, Wang L, Cai Q, Wang Y
Association Between Cancer Incidence and Mortality in Web-Based Data in China: Infodemiology Study
J Med Internet Res 2019;21(1):e10677
URL: https://www.jmir.org/2019/1/e10677/
doi:10.2196/10677
PMID:30694203

Review

Evaluating Digital Maturity and Patient Acceptability of Real-Time Patient Experience Feedback Systems: Systematic Review

Mustafa Khanbhai[1], MBChB, BScHons, MRCS (Ed); Kelsey Flott[1], BA, MSc; Ara Darzi[1], PC, KBE, FRS, FMedSci, HonFREng; Erik Mayer[1], MBBS, BSc, MRCS, PhD, FRCS (Urol)

Centre for Health Policy, Imperial College London, London, United Kingdom

Corresponding Author:
Mustafa Khanbhai, MBChB, BScHons, MRCS (Ed)
Centre for Health Policy
Imperial College London
10th Floor QEQM Building
Centre for Health Policy
London, W2 1NY
United Kingdom
Phone: 44 02033126428
Fax: 44 207594184
Email: m.khanbhai@imperial.ac.uk

Abstract

Background: One of the essential elements of a strategic approach to improving patients' experience is to measure and report on patients' experiences in real time. Real-time feedback (RTF) is increasingly being collected using digital technology; however, there are several factors that may influence the success of the digital system.

Objective: The aim of this review was to evaluate the digital maturity and patient acceptability of real-time patient experience feedback systems.

Methods: We systematically searched the following databases to identify papers that used digital systems to collect RTF: The Cochrane Library, Global Health, Health Management Information Consortium, Medical Literature Analysis and Retrieval System Online, EMBASE, PsycINFO, Web of Science, and CINAHL. In addition, Google Scholar and gray literature were utilized. Studies were assessed on their digital maturity using a Digital Maturity Framework on the basis of the following 4 domains: capacity/resource, usage, interoperability, and impact. A total score of 4 indicated the highest level of digital maturity.

Results: RTF was collected primarily using touchscreens, tablets, and Web-based platforms. Implementation of digital systems showed acceptable response rates and generally positive views from patients and staff. Patient demographics according to RTF responses varied. An overrepresentation existed in females with a white predominance and in patients aged ≥65 years. Of 13 eligible studies, none had digital systems that were deemed to be of the highest level of maturity. Three studies received a score of 3, 2, and 1, respectively. Four studies scored 0 points. While 7 studies demonstrated capacity/resource, 8 demonstrated impact. None of the studies demonstrated interoperability in their digital systems.

Conclusions: Patients and staff alike are willing to engage in RTF delivered using digital technology, thereby disrupting previous paper-based feedback. However, a lack of emphasis on digital maturity may lead to ineffective RTF, thwarting improvement efforts. Therefore, given the potential benefits of RTF, health care services should ensure that their digital systems deliver across the digital maturity continuum.

(J Med Internet Res 2019;21(1):e9076) doi:10.2196/jmir.9076

KEYWORDS

digital maturity; digital technology; feedback; patient experience; real time

Introduction

Background of Patient Experience Feedback

Alongside measures of clinical effectiveness and safety outcomes, patient experience is increasingly recognized as an important indicator of the quality of health care provision and is frequently cited in health policies [1-3]. Yet, in practice and research, the concept of patient experience has had varied uses and is often discussed with little more explanation than the term itself [4]. In 2011, the English National Health Service (NHS) [5] outlined 8 domains that help define patient experience and are critical to patients' experience of health care services. This has been used as an agreed working definition of patient experience to guide the measurement of patient experience across the NHS.

Patient Experience Feedback in Real Time

One of the essential elements of a strategic approach to improving patients' experience is to measure and report on patients' experiences to assess progress, strengthen accountability, and identify new opportunities for improving performance [6]. Evidence suggests that this can be achieved using real-time feedback (RTF) [7]. RTF involves the systematic collection, analysis, and reporting of information from individuals at the point of care [8]. Previous studies have found that RTF has the potential to enable health care organizations to respond promptly to patients' concerns and make timely improvements to services [7,9-11].

Collecting Real-Time Feedback Using Digital Technology

With the ability to maximize the scalability and speed of data collection while reducing cost [12], digital tools (tablets, kiosks, emailed survey, and websites) are increasingly being utilized to gather patient experience feedback [13,14], allowing summary results to be reported on an ongoing [15] real-time basis. In a recent Cochrane review [16], self-administered survey questionnaire responses collected using mobile apps were compared with those collected using other methods, and it was concluded that the delivery of survey questionnaires through apps does not affect data equivalence and can improve data completeness. However, none of the questionnaires evaluated patient experience.

There is a growing pressure on health care services to embrace digital technologies tosignificantly improve the patient experience [17]. With an increase in adoption of digital systems pertaining to RTF, health care services must recognize and overcome the barriers that may hinder successful integration and uptake of such technologies. One of the ways of achieving this is through systematic evaluation and monitoring of digital systems to ensure they operate in the way they are intended and cultivate a better patient experience [17]. Digital maturity—the extent to which digital technologies are used as enablers to deliver a high-quality health service—is an emerging concept across developed health care systems [17]. Evidence suggests that digital maturity is linked to better outcomes and is indicative of a well-performing organization [18]. Evaluating digital systems for digital maturity highlights where gaps exist in maturity, which presents an opportunity to address the specific shortcomings [17]. The digital maturity of the systems that are being used to collect RTF has not been previously evaluated. A deeper understanding of the limitations of individual digital systems may help health care services pinpoint areas in need of improvement and organizational change. Without this, digital systems used for RTF may not be successful, and patients' voices may go unheard.

Therefore, this systematic review aims to evaluate the digital maturity of digital RTF systems. Specific objectives were to (1) describe the digital modes utilized; (2) provide insights into patients' and staff views of these digital systems; and (3) demonstrate digital maturity by systematically scoring individual digital systems.

Methods

Search Strategy

The following databases were searched: Medical Literature Analysis and Retrieval System Online, EMBASE, PsycINFO, The Cochrane Library (Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register), Global Health, Health Management Information Consortium, CINAHL, and Web of Science. In addition, gray literature and Google Scholar were utilized to extract papers that were not retrieved in the databases searched. Publications from January 2000 to February 2017 were included. We limited our electronic searches to studies on or after 2000 because it was around this time when the digital technology revolution in health care began to emerge [19]. Owing to the diversity of terms used inferring patient experience, combinations of search terms were used. Multimedia Appendix 1 provides the complete list of subheadings (Medical Subject Headings) and keywords.

Inclusion Criteria

Studies were deemed eligible for inclusion if (1) digital modes of administration were employed; (2) collection was in real time (at the point of care) or near real time (immediately after discharge); and (3) they were conducted in a primary or secondary care setting. There was no restriction on the age of the patient population of interest. Both quantitative and qualitative studies were included. Multimedia Appendix 1 provides further details of the inclusion criteria.

Search Flow

The research adhered to the guideline presented in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2009 checklist [20]. Data analysis involved the comparison of included studies and extracted data. Due to the heterogeneous nature of the studies, a narrative synthesis was deemed most appropriate. This was followed by scoring the digital systems reported in the studies to determine the digital maturity.

Digital Maturity

An existing framework [17] was utilized to determine the digital maturity of individual digital systems. This framework is embedded in the literature and has the benefit of systematically

assessing the effectiveness of any digital system in health care. Each included study was scored across four key domains—capacity or resource, usage, interoperability, and impact. The framework highlights key questions for each domain and is based on "yes" or "no" response to each question. We assigned a maximum of 1 point if the digital system in each study demonstrated appropriate evidence for that particular domain. The maximum overall points a study could achieve was 4. This indicates the overall success of the digital platform [17]. Two reviewers (MK and KF) independently scored each study; disagreements in scoring were resolved by discussion between the two reviewers. Interrater agreement *kappa* was calculated.

Results

Overall Description

The initial search returned 3456 papers; after removing duplicates, 3438 papers were retained. The titles and abstracts were screened, and 112 papers were identified as potentially eligible for inclusion. Full-text papers were retrieved and assessed for inclusion, of which 13 were retained for final inclusion. RTF was defined as feedback collected while patients are in a hospital, at the point of care, while receiving care, or immediately after discharge. The main reason for exclusion was the papers did not report patient experiences or reported user experience of other digital technology. Figure 1 illustrates the PRISMA flowchart representing the study selection process and reasons for exclusion.

The reasons for exclusion of records (n=3326) and full-text papers (n=99) were the following: nondigital mode of administration (n=241), not real time or near real time (n=130), assessment or evaluation of other parameters (n=1489), feasibility or usability testing of digital systems not related to patient experience data collection (n=655), experience of other digital technology (n=848), and reviews (n=62).

Systematic Review of Studies

Study Characteristics

Of the 13 studies included in the final review (Table 1), 5 were based in general practice [10,21-24] and the rest were based in an acute care hospital setting [9,25-31]. Of all the studies, 5 were from the United Kingdom [21,22,24,25,28], 7 from the United States [9,10,26,27,29-31], and 1 from Canada [23]. All but one [28] studies were based on adult populations. However, the experience was reported by adults (parents or relatives) for the study [28] conducted in the neonatal ward. Five studies were qualitative studies on patients' or staff views [21,24] and on barriers to and facilitators [25,27,28] of RTF. The remainder of the studies used a quantitative approach with varied outcome measures, of which the recurring measures were response rates of RTF [9,10,22,23,27,29] and association with patient demographics, that is, age, gender, ethnicity, and literacy [9,22,23,31].

Figure 1. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2009 checklist.

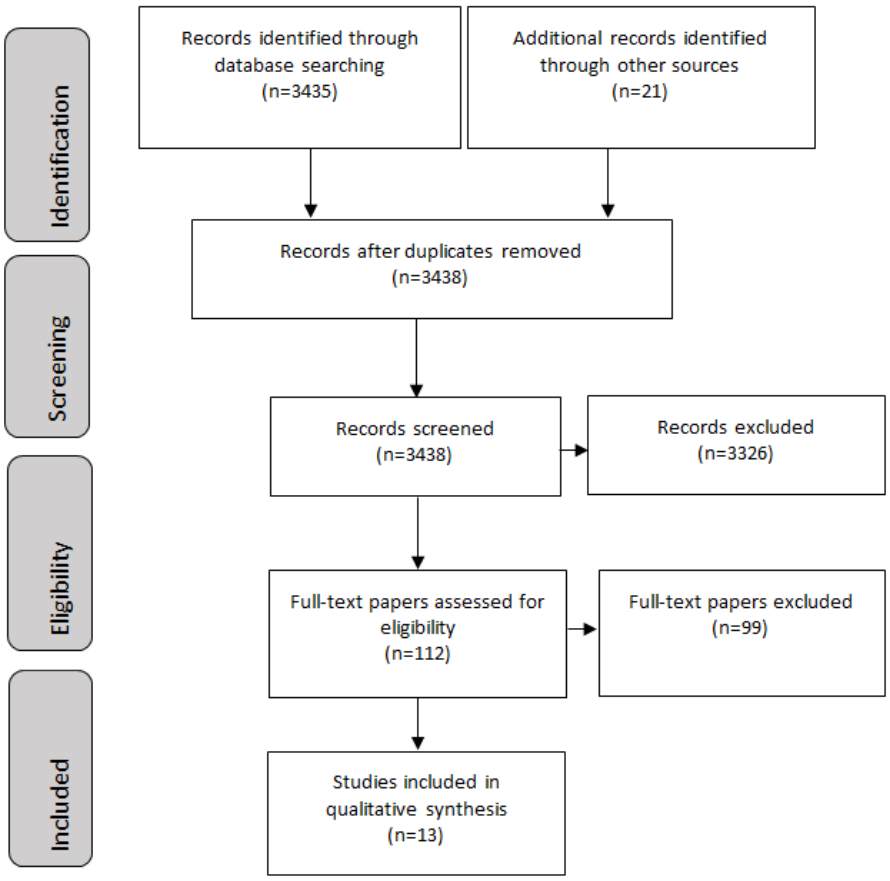


Table 1. Study characteristics of the 13 studies included in the systematic review.

Publication title; author(s), year	Study design	Duration of study	Types of survey questionnaire(s)	Mode of administration
Capturing patient experience: a qualitative study implementing real-time feedback in primary care; Carter M et al, 2016 [21]	Qualitative	3 months	Modified Staff opinions (semistructured interviews and focus groups) using Normalization Process Theory	Kiosks
Patients' use and views of real-time feedback technology in general practice; Wright C et al, 2017	Exploratory randomized trial	3 months	Modified (amalgamated Friends and Family Test, 6 items focusing on access, communication and satisfaction [derived from general practitioner patient survey], 2 practices tailored questions[a])	2 touch screens (1 Kiosk and 1 desktop device)
Measuring the patient experience in primary care. Comparing e-mail and waiting room survey delivery in a family health team; Slater M et al, 2016 [23]	Cross-sectional comparative analysis	1 month	Modified (amalgamated Commonwealth Fund International Health Policy Survey, patient demographics, self-rated health)	Tablet
Barriers and facilitators of a near real-time feedback approach for measuring patient experiences of hospital care; Kasbauer et al, 2017 [25]	Qualitative	10 months	Novel and validated (Compassionate Care Toolkit)	Tablet and kiosks
Real-time patient survey data during routine clinical activities for rapid-cycle quality improvement; Wofford et al, 2015 [10]	Feasibility study	1 month	Modified (dental care, waiting room experience, continuity, and internet access)	Tablet
Real-time patient experience surveys of hospitalized medical patients; Indovina et al, 2016 [26]	Prospective randomized	5 months	Previously validated (US Department of Health and Human Service and the Hospital Consumer Assessment of Healthcare Providers and Systems [HCAHPS])	Web-based platform
Evaluating patient-centred care (PCC): pilot study testing feasibility of electronic data collection in hospitalised older adults; Duffy et al, 2012 [27]	Cross-sectional feasibility	3 months	Previously validated (e-Caring Assessment Tool)	Tablet
Patient experience tracker (PET) survey as measure of quality in the neonatal unit; Aladangady et al, 2011 [28]	Qualitative	N/A[b]	Novel	Tablet
Development and validation of the tool to assess inpatient satisfaction with care from hospitalists; Torok et al, 2014 [29]	Cross-sectional	3 months	Novel and validated (Tool to Assess Inpatient Satisfaction with Care from Hospitalists)	Tablet Paper
Incentivized digital outcomes collection; Isenberg S et al, 2001 [30]	Feasibility study	3 months	Previously validated (Visit Rating Questionnaire)	Telephone (electronic voice response technology)
Exploring patients' views toward giving Web-based feedback and rating to general practitioners in England: a qualitative descriptive study; Patel et al, 2016 [24]	Descriptive exploratory qualitative approach	N/A	Previously validated (Friends and Family Test)	Web-based platform (National Health Service Choices) Paper
Obtaining patient feedback at point of service using electronic kiosks; Dirocco et al, 2011 [9]	Feasibility study	1.5 months	Modified (National Committee for Quality Assurance's Healthcare Effectiveness Data and Information Set and Quality Measurement standards)	Kiosk
Improving patient satisfaction through physician education, feedback, and incentives; Banka et al, 2015 [31]	Nonrandomized comparative study	14 months (7 months each year)	Previously validated (Assessing Residents' Connect with patients, Introduce yourself and role, Communicate, Ask and anticipate, Respond, and Exit courteously and HCAHPS)	Paper and Web-based platform (results sent via email)

[a]Filter questions (tailored to patients visit).

[b]N/A: not applicable.

Modes of Feedback

RTF was collected using touchscreens (kiosk) [9,21,22,25], tablets [10,23,25,27-29], and Web-based platform [24,26,30,31] in the included studies. With regards to the timing of feedback, 11 studies collected their feedback in real time [9,10,21-23,25,27-31] and 2 studies in near real time [24,30] (ie, within 48 hours of discharge). Patient experience questionnaires used in each study varied; 5 studies modified an existing questionnaire [9,10,21-23], 5 used previously validated questionnaires [24,26,27,30,31], 2 used novel and validated questionnaires [25,29], and 1 used a novel and nonvalidated questionnaire [28]. The majority of the questionnaires were in English only.

Response Rates of Real-Time Feedback

Of the 13 studies, 6 studies [9,10,22,23,27,29] evaluated response rates (percentage) as part of their outcome measures. The response rates were 55.9% [23], 43.4% [10], and 2.5% [22] for studies conducted in primary care and 54.9% [9], 59.2% [27], and 61.5% [29] for those conducted in secondary care. Multimedia Appendix 2 demonstrates the absolute numbers of responses in each study identified above. Only 1 study evaluated response rates of RTF compared with non-RTF and showed that RTF improved response rates (55.9% vs 19.8%) [23].

Completion Time

Of all studies, 4 evaluated time to completion, and this varied from 40.4 seconds [10], <2 minutes [22], 3 minutes [9] to 31 minutes [27]. However, each questionnaire was different as it varied in length and number of questions.

Patient Demographics According to Real-Time Feedback Responses

The patient demographics that were collected varied, and 7 [9,22,23,25,27,29,31] of the 13 studies evaluated responses by patient demographics. Only Slater et al [23] compared RTF and non-RTF responses by patient demographics; RTF showed a higher percentage response in males (43.5 vs 36.6), in age (years) bracket 18-24 and 24-34 (6.7 vs 3.6 and 23.8 vs 18.5, respectively), and in those with lower literacy (34.4 vs 21.3) compared with non-RTF. Banka et al [31] revealed a higher percentage of male respondents (55.3 vs 41.4) with a white predominance (62.5 vs 60.9) using RTF compared with non-RTF. However, Wright et al [22] revealed different findings. There was a higher percentage response in females in the 46-65 (26.2%) and >65 (33.8%) age range and higher response from white than from ethnic patients. Dirocco et al [9] showed mixed results in that although the response was higher from females, there was a higher percentage response from African American individuals than from white American individuals (48.5 vs 38.6) and those aged 18-49 years (45.2). Torok et al [29] and Duffy et al [27] showed that the response rates from elderly patients were adequate; however, of these elderly patients, there was an overrepresentation in females [27,29], white individuals [27,29], and educated patients [27] who were aged ≥60 years. Kasbauer et al [25] showed that elderly patients responded to the survey but with the help of volunteers. Multimedia Appendix 3 summarizes the findings from the included studies.

Patient Views of Real-Time Feedback

Patients' views of RTF were collected in 6 studies [9,10,22,24,27,30]. The main pros of RTF were as follows: easy to complete [10,22], easy to use [9,27], and willingness to use [24,27] the data collection tool. Isenberg et al [30] showed that patients were motivated to use RTF; however, this was incentivized with the reward of free long-distance minutes for patients and with a practice improvement program for the staff. Patel et al [24] identified that younger patients (aged <50 years) found digital platforms more accessible compared with older ones (aged ≥60 years). However, Duffy et al [27] further evaluated opinions of older adults and found a preference toward digital platforms; for example, 70% of patients preferred to answer questions using an iPad. Four studies found that patients thought RTF completion was quick with fast turnaround time [9,10,22,27]. The main cons of RTF were as follows: lack of awareness of the opportunity to leave feedback [22,24], lack of time [22], concerns over technology [22,24], concerns over anonymity [22,24], and age- or disease-related exclusion [24,27]. Interestingly, Wright et al [22] showed that those who did not use RTF were still positive about the idea of providing RTF.

Staff Views of Real-Time Feedback

Staff views were collected in 4 studies [9,21,25,28]. Positive staff views toward RTF were as follows: immediacy of RTF compared with traditional surveys, which helped offset "feedback fatigue," complemented other forms of feedback with the potential of integrating with other data sources [21]. Staff found the RTF data was felt to be useful when summarized, highlighting areas of improvement at-a-glance on a dashboard [25], and when there was coworking with senior clinical staff [25], increasing staff morale and awareness of RTF. In some studies, free text was found to be more useful than quantitative questions [21,25] as it brought the experiences to life for frontline staff and added a "sense of urgency" to address them in improvement efforts [25]. Concerns included duplication with other forms of feedback, lack of time for patients to reflect on experience, extreme views, exclusion of certain patient groups, staff did not feel included in decision making [21], initial reluctance [28], limited time to review, and lack of access to the results [25].

Evaluating Digital Maturity of Real-Time Feedback Systems

Three studies received a score of 3 [10,25,30], 2 [9,21,22], or 1 [27,28,31] points, respectively, while 4 studies [23,24,26,29] were attributed 0 points. While 7 studies demonstrated capacity or resource, 8 demonstrated impact. None of the studies demonstrated digital systems that were deemed to be mature, that is, did not achieve a full score in all of the 4 domains. We describe in detail how the digital system in each study demonstrated evidence that determined whether a point was given in each of the 4 domains. Following independent scoring by MK and KF, Cohen *kappa* was calculated as 0.98, suggesting an almost perfect agreement. There was only one domain where the scoring differed (Usage) [26], and through discussion, a final score of 0 was inputted. Multimedia Appendix 4 details the individual scoring with a description on each domain.

Discussion

Principal Findings

This review highlights that digital modes of administration of RTF are well accepted by patients and staff, with response rates equivalent to, and in some studies better than, non-RTF. From a patient's perspective, it has been reported that its influx alongside demographic shifts such as increasing aging population and ethnicity [32,33] can lead to patient disengagement and poor uptake of these technologies [34]. On the contrary, we have shown that patients are in fact willing to engage with this technology for experience reporting, thereby disrupting previous nonreal-time, paper-based feedback. However, digital technology may not be preferred by all patients; therefore, health care organizations need to be mindful and consider other means of inclusivity when developing digital health care systems. From a staff perspective, although RTF is well received, problems arising from the lack of robust digital infrastructure [25] can thwart improvement efforts. From a health care organizational perspective, most digital systems were unable to demonstrate interoperability and very few demonstrated impact and, therefore, were not deemed digitally mature, compromising success within the organization.

Digital Maturity of Existing Real-Time Feedback

By nature, maturity frameworks not only identify components of a successful system but also capture the evolution of a digital system from conception to implementation to impact. Using evidence where possible from the included studies, we highlight how each of the 4 domains contribute to ensuring digital maturity.

Capacity or Resource

The success of digital health is contingent on establishing the necessary capacity and resources to build, use, and support access to high-quality health services and harvest useful information in the health system. There was a general lack of analytical support in most digital systems to extract valuable information such as the ability to examine user-specific interaction [10] or adjust survey responses [23]. A Kings' Fund report explains that gleaning information from experience data requires the same analytical capability as interpreting clinical data; however, this is often unavailable [35]. The resource capacity requirements stretch beyond health professionals and technology specialists across the continuum of care to include health information managers to information security professionals [36]. Staff time was an important barrier in data collection, and this is in keeping with other studies that reported a lack of time or resources to collect, analyze, or act on data and need for staff training in data analysis and statistics to facilitate full understanding and use of results [37-39]. To circumvent this, some of the studies used volunteers [25,31] and incentives [30,31] to gather data. This generates concerns such as response bias and competition among clinicians, and there may be an element of the Hawthorne effect, whereby clinicians modify their behavior as they are being observed, which may explain the positive outcomes in those studies. Going forward, a strategic approach should take into consideration building human and institutional capacity, nurturing clinical and community champions, and developing the base of knowledgeable users to drive appropriate adoption of digital systems [36].

Usage

As the needs and experiences differ greatly across these groups, a flexible or responsive data collection mode is needed, which can aid patients during the data collection process. While most patients in the included studies were comfortable using the digital system and required little prompting or help once they engaged with the survey, some differences were noted. Concerns exist that older patients are less comfortable with technology [40]; however, there was an overrepresentation in responses from patients aged >65 years in the included studies [22,25,27,29,31]. Moreover, some of these studies were conducted in the elderly population [25,27]. Despite this, the response rate in this patient group was adequate, suggesting that their movement into digital life is evolving. Furthermore, when certain conditions are met such as providing a supportive, nonhurried environment [41]; bold, plain, and large font with fewer graphics; and avoidance of certain colors [42], older adults can be successful users. The use of videos in the software can enhance accuracy and acceptability [10]. Furthermore, trained volunteers can provide a responsive approach to real-time data collection from lesser heard groups [14], increasing patient engagement and, subsequently, improving response rates. In addition, they can help reduce the data collection burden, which may otherwise fall on clinical staff and may even account for the false positives as seen in some surveys [43] due to the presence of staff during survey completion.

If data collection is obtrusive, unrealistic, or inaccurate, it can undermine enthusiasm for assessing and improving practice quality [44,45]. However, the digital systems in this review demonstrated quick turnaround of data collection, collation, and dissemination of results. This was key to successful implementation of digital systems as it promoted "buy-in" from staff. Furthermore, patients are fatigued of requests for feedback in health care and in daily life; hence, data collection should be quick, focused, and part of routine care to encourage participation from patients and clinicians and to be sustainable in a busy setting [10].

Interoperability

From a health care perspective, interoperability is needed to reform the chaotic, and, at times, dysfunctional nature of how information is shared within and among health care services. There was a ubiquitous shortfall in achieving interoperability among the digital systems in this review. Data that are not interoperable cannot be analyzed alongside other data indicative of care quality. This perpetuates the siloed approach to data interpretation and creates a chasm between data and information for improvement. Some of the challenges in achieving interoperability include resistance from some vendors, prohibitively high data exchange fees, and lack of incentives to develop interoperability and technical variations. Without tackling the interoperability of RTF systems, it is likely that health care organizations will fail on delivering impactful quality improvement. This must now be a major focus for health care services, but it should be done in an organized way that

prioritizes interoperability so that patient feedback data can communicate seamlessly to generate information that will benefit health care services and patients alike.

Impact

To genuinely demonstrate impact, the digital system should not only be able to generate quality improvement activities but also demonstrate sustainability and cost-effectiveness. Three studies [10,25,30] achieved a score for impact as they demonstrated a change in practice following the implementation of RTF. However, the majority of the digital systems were not able to demonstrate any impact. The duration of studies in this review was short (between 1 and 12 months), and the sustainability of the digital system to continue to deliver quality improvement could not be evaluated. Without considerations of sustainability, digital programs are of limited value.

Limitations

Due to publication bias, there may be unpublished evidence of nonsignificant or negative findings or findings held locally, which are not published or otherwise publicly available. We ascribed a level of digital maturity of individual systems based on the evidence provided in the studies. If this was not discernible while reviewing the studies, we assumed it to be lacking. As the remit of the included papers differed greatly, their digital systems may, in fact, have existing procedures for ensuring capacity, usage, interoperability, and impact, but the authors did not specify this information in the published paper. Of note, the studies were not evaluated for research quality before being included in the synthesis owing to the limited number of studies that were identified after exclusion.

Conclusions

There was not a large body of published literature on digital modes in relation to RTF. However, the evidence provided by the studies in this review demonstrates that there is a potential in using digital modes of administration of RTF as an agent for improving service delivery. Patients and staff alike are willing to engage in RTF, demonstrated by acceptable response rates. However, for RTF to be impactful, health care organizations must ensure that they have strategies in place to deliver across all levels of digital maturity. Health care services have the capacity to introduce digital solutions for RTF; however, lack of interoperability is slowing down progress. In addition, it is possible that some health care services may be wasting effort and resources when they invest in digital technologies for RTF. On balance, the direction of the health care ecosystem toward embracing digital technology looks promising, and as health care shifts toward a patient-centered model, digital technology will be an important partner in this transformation.

Acknowledgments

The research was supported by the National Institute for Health Research (NIHR) Imperial Patient Safety Translational Research Centre, NIHR Imperial Biomedical Research Centre, and The Health Foundation. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health.

Authors' Contributions

MK, KF, and EM contributed to the design of the study. MK and KF managed the review process. MK led on screening, data extraction, and scoring. KF assisted in data synthesis and scoring. MK drafted the manuscript, and KF, AD, and EM reviewed the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Subheadings (MESH) and Keywords.

[PDF File (Adobe PDF File), 35KB - jmir_v21i1e9076_app1.pdf]

Multimedia Appendix 2

This demonstrates the percentage response rates for RTF completion including absolute number of patients in each study.

[PDF File (Adobe PDF File), 16KB - jmir_v21i1e9076_app2.pdf]

Multimedia Appendix 3

Response rate in percentage and representation of responses according to patients' demographics documented from the studies in the systematic review.

[PDF File (Adobe PDF File), 50KB - jmir_v21i1e9076_app3.pdf]

Multimedia Appendix 4

Individual scores in relation to the digital maturity framework domains. A score of one indicates that the study demonstrates evidence in that particular domain, otherwise a score of zero is ascribed.

[PDF File (Adobe PDF File), 50KB - jmir_v21i1e9076_app4.pdf]

References

1. Darzi A. High quality care for all: NHS Next Stage Review final report. 2008. Department of Health URL: https://www.gov.uk/government/publications/high-quality-care-for-all-nhs-next-stage-review-final-report [accessed 2018-11-15] [WebCite Cache ID 73wwxotdd]
2. Department of Health and Social Care. Gov.UK. 2010 Jul 12. Liberating the NHS white paper URL: https://www.gov.uk/government/publications/liberating-the-nhs-white-paper [accessed 2018-11-15] [WebCite Cache ID 73wyJ0gAy]
3. Foot C, Fitzsimons B. The policy and practice of measuring patient experience. Nurs Manag (Harrow) 2011 Jun;18(3):18-19. [doi: 10.7748/nm2011.06.18.3.18.c8535] [Medline: 21736168]
4. Niederhauser V, Wolf J. Patient Experience: A Call to Action for Nurse Leadership. Nurs Adm Q 2018;42(3):211-216. [doi: 10.1097/NAQ.0000000000000293] [Medline: 29870486]
5. NHS National Quality Board. Department of Health. 2011. NHS Patient Experience Framework URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215159/dh_132788.pdf [accessed 2018-11-15] [WebCite Cache ID 73wyKDC71]
6. Coulter A. The King's Fund. 2012. Leadership for patient engagement URL: https://www.kingsfund.org.uk/sites/default/files/leadership-patient-engagement-angela-coulter-leadership-review2012-paper.pdf [accessed 2018-11-15] [WebCite Cache ID 73wyPmMRT]
7. Northumbria Healthcare NHS Foundation Trust. 2017 Apr 13. Northumbria supports patient experience feedback across the NHS URL: https://www.northumbria.nhs.uk/northumbria-supports-patient-experience-feedback-across-nhs/ [accessed 2017-12-01]
8. Garrisson M, Wolf J. The Beryl Institute. 2016 Apr. The Role of the Volunteer in Improving Patient Experience URL: http://www.theberylinstitute.org/news/284517/The-Role-of-the-Volunteer-in-improving-Patient-Experience-Explored-by-The-Beryl-Institute-.htm [accessed 2017-11-12]
9. Dirocco DN, Day SC. Obtaining patient feedback at point of service using electronic kiosks. Am J Manag Care 2011 Jul 01;17(7):e270-e276 [FREE Full text] [Medline: 21819174]
10. Wofford JL, Campos CL, Jones RE, Stevens SF. Real-time patient survey data during routine clinical activities for rapid-cycle quality improvement. JMIR Med Inform 2015 Mar 12;3(1):e13 [FREE Full text] [doi: 10.2196/medinform.3697] [Medline: 25768807]
11. Larsen D, Peters H, Keast J, Devon R. Using real time patient feedback to introduce safety changes. Nurs Manag (Harrow) 2011 Oct;18(6):27-31. [doi: 10.7748/nm2011.10.18.6.27.c8718] [Medline: 22017150]
12. Greenlaw C, Brown-Welty S. A comparison of web-based and paper-based survey methods: testing assumptions of survey mode and response cost. Eval Rev 2009 Oct;33(5):464-480. [doi: 10.1177/0193841X09340214] [Medline: 19605623]
13. Lane SJ, Heddle NM, Arnold E, Walker I. A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. BMC Med Inform Decis Mak 2006 May 31;6:23 [FREE Full text] [doi: 10.1186/1472-6947-6-23] [Medline: 16737535]
14. Shih T, Fan X. Comparing response rates in e-mail and paper surveys: A meta-analysis. Educational Research Review 2009 Jan;4(1):26-40. [doi: 10.1016/j.edurev.2008.01.003]
15. Kelsey T. NHS England. 2014 Jul. The Friends and Family Test URL: https://www.england.nhs.uk/wp-content/uploads/2014/07/fft-imp-guid-14.pdf [accessed 2018-11-15] [WebCite Cache ID 73wzhs2za]
16. Marcano Belisario JS, Jamsek J, Huckvale K, O'Donoghue J, Morrison CP, Car J. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. Cochrane Database Syst Rev 2015 Jul 27(7):MR000042. [doi: 10.1002/14651858.MR000042.pub2] [Medline: 26212714]
17. Flott K, Callahan R, Darzi A, Mayer E. A Patient-Centered Framework for Evaluating Digital Maturity of Health Services: A Systematic Review. J Med Internet Res 2016 Apr 14;18(4):e75 [FREE Full text] [doi: 10.2196/jmir.5047] [Medline: 27080852]
18. NHS England Data Catalogue. Digital Maturity Assessment 2015/2016 URL: https://data.england.nhs.uk/dataset/digital-maturity-assessment-2015-2016/resource/06f03a99-6451-4c15-a795-99aafc840cf0 [accessed 2018-11-15] [WebCite Cache ID 73x08kGa4]
19. Macnaughtan L. Innovatemedtec. 2015 May 25. The Curious Case of Digital Health URL: https://innovatemedtec.com/content/the-curious-case-of-digital-health [accessed 2017-11-13]
20. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

21. Carter M, Davey A, Wright C, Elmore N, Newbould J, Roland M, et al. Capturing patient experience: a qualitative study of implementing real-time feedback in primary care. Br J Gen Pract 2016 Nov;66(652):e786-e793 [FREE Full text] [doi: 10.3399/bjgp16X687085] [Medline: 27621292]
22. Wright C, Davey A, Elmore N, Carter M, Mounce L, Wilson E, et al. Patients' use and views of real-time feedback technology in general practice. Health Expect 2017 Dec;20(3):419-433 [FREE Full text] [doi: 10.1111/hex.12469] [Medline: 27124589]
23. Slater M, Kiran T. Measuring the patient experience in primary care: Comparing e-mail and waiting room survey delivery in a family health team. Can Fam Physician 2016 Dec;62(12):e740-e748 [FREE Full text] [Medline: 27965350]
24. 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 Dec 05;18(8):e217 [FREE Full text] [doi: 10.2196/jmir.5865] [Medline: 27496366]
25. Käsbauer S, Cooper R, Kelly L, King J. Barriers and facilitators of a near real-time feedback approach for measuring patient experiences of hospital care. Health Policy Technol 2017 Mar;6(1):51-58 [FREE Full text] [doi: 10.1016/j.hlpt.2016.09.003] [Medline: 28367401]
26. Indovina K, Keniston A, Reid M, Sachs K, Zheng C, Tong A, et al. Real-time patient experience surveys of hospitalized medical patients. J Hosp Med 2016 Apr;11(4):251-256. [doi: 10.1002/jhm.2533] [Medline: 26777721]
27. Duffy JR, Kooken WC, Wolverton CL, Weaver MT. Evaluating patient-centered care: feasibility of electronic data collection in hospitalized older adults. J Nurs Care Qual 2012;27(4):307-315 [FREE Full text] [doi: 10.1097/NCQ.0b013e31825ba9d4] [Medline: 22677911]
28. Aladangady N, Negus J. Patient Experience Tracker (PET) survey as measure of quality in the Neonatal Unit. Clinical Risk 2011 May 24;17(3):88-91. [doi: 10.1258/cr.2011.010086]
29. Torok H, Ghazarian SR, Kotwal S, Landis R, Wright S, Howell E. Development and validation of the tool to assess inpatient satisfaction with care from hospitalists. J Hosp Med 2014 Sep;9(9):553-558. [doi: 10.1002/jhm.2220] [Medline: 24888242]
30. Isenberg SF, Davis CL, Adams CE, Isenberg JS, Isenberg MA. Incentivized digital outcomes collection. Am J Med Qual 2001;16(6):202-211. [doi: 10.1177/106286060101600604] [Medline: 11816851]
31. Banka G, Edgington S, Kyulo N, Padilla T, Mosley V, Afsarmanesh N, et al. Improving patient satisfaction through physician education, feedback, and incentives. J Hosp Med 2015 Aug;10(8):497-502. [doi: 10.1002/jhm.2373] [Medline: 26014339]
32. Goldzweig C, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
33. 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] [Medline: 26104044]
34. Krane D, Vella A, Bechtel C, Lederer L, Padget J, Latham B, et al. National Partnership for Women & Families. 2014 Dec. Engaging Patients and Families: How Consumers Value and Use Health IT URL: http://www.nationalpartnership.org/research-library/health-care/HIT/engaging-patients-and-families.pdf [accessed 2018-11-15] [WebCite Cache ID 73x1eFJr4]
35. Robert GC, Cornwell J, Brearley S, Foot C, Goodrich J, Joule N, et al. Developing the evidence base for measuring and improving patient experience. NHS Institute for Innovation and Improvement 2011 Nov;London:1-200 [FREE Full text]
36. Ho K, Al-Shorjabji N, Brown E, Zelmer J, Gabor N, Maeder A, et al. Applying the Resilient Health System Framework for Universal Health Coverage. Stud Health Technol Inform 2016;231:54-62. [Medline: 27782016]
37. Barr JK, Giannotti TE, Sofaer S, Duquette CE, Waters WJ, Petrillo MK. Using public reports of patient satisfaction for hospital quality improvement. Health Serv Res 2006 Jun;41(3 Pt 1):663-682 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00508.x] [Medline: 16704506]
38. Reeves R, Seccombe I. Do patient surveys work? The influence of a national survey programme on local quality-improvement initiatives. Qual Saf Health Care 2008 Dec;17(6):437-441 [FREE Full text] [doi: 10.1136/qshc.2007.022749] [Medline: 19064659]
39. Davies E, Shaller D, Edgman-Levitan S, Safran DG, Oftedahl G, Sakowski J, et al. Evaluating the use of a modified CAHPS survey to support improvements in patient-centred care: lessons from a quality improvement collaborative. Health Expect 2008 Jun;11(2):160-176 [FREE Full text] [doi: 10.1111/j.1369-7625.2007.00483.x] [Medline: 18494960]
40. Huntington P, Williams P, Nicholas D. Age and gender differences of a touch-screen kiosk: a study of kiosk transaction log files. Inform Prim Care 2002 Mar;10(1):3-9 [FREE Full text]
41. Murata A, Iwase H. Usability of touch-panel interfaces for older adults. Hum Factors 2005;47(4):767-776. [doi: 10.1518/001872005775570952] [Medline: 16553065]
42. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
43. Insight Team, NHS England. NHS England. 2014 Jul 21. Review of the Friends and Family Test URL: https://www.england.nhs.uk/wp-content/uploads/2014/07/fft-rev1.pdf [accessed 2018-11-15] [WebCite Cache ID 73x2PqiaN]
44. Baxley EG, Bennett KJ, Pumkan C, Crutcher S, Helms MG. “PDSA-ADHD”: a newly reported syndrome. J Am Board Fam Med 2011;24(6):752-757.

45. Price EL, Mackenzie TD, Metlay JP, Camargo CA, Gonzales R. A computerized education module improves patient knowledge and attitudes about appropriate antibiotic use for acute respiratory tract infections. Patient Educ Couns 2011 Dec;85(3):493-498. [doi: 10.1016/j.pec.2011.02.005] [Medline: 21392929]

Abbreviations

NHS: National Health Service
NIHR: National Institute for Health Research
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RTF: real-time feedback

Edited by G Eysenbach; submitted 28.09.17; peer-reviewed by H Brown, L Kayser, L Wintner, N Ribeiro, K Agarwal, D Nault; comments to author 09.12.17; revised version received 24.04.18; accepted 24.09.18; published 14.01.19.

Please cite as:
Khanbhai M, Flott K, Darzi A, Mayer E
Evaluating Digital Maturity and Patient Acceptability of Real-Time Patient Experience Feedback Systems: Systematic Review
J Med Internet Res 2019;21(1):e9076
URL: https://www.jmir.org/2019/1/e9076/
doi:10.2196/jmir.9076
PMID:31344680

Original Paper

Additional Telemedicine Rounds as a Successful Performance-Improvement Strategy for Sepsis Management: Observational Multicenter Study

Robert Deisz[1*], MD; Susanne Rademacher[1*], MSc; Katrin Gilger[1], MD; Rudolf Jegen[2], MD; Barbara Sauerzapfe[3], MD; Christina Fitzner[4], Dipl stat; Christian Stoppe[1], PhD, MD; Carina Benstoem[1], MSc, PhD; Gernot Marx[1], MD, FRCA, PhD

[1]Department of Intensive Care Medicine, Medical Faculty RWTH Aachen, University Hospital RWTH Aachen, Aachen, Germany

[2]Department of Anaesthesiology, St. Elisabeth Hospital, Jülich, Germany

[3]Department of Anaesthesiology, Franziskushospital, Aachen, Germany

[4]Department of Medical Statistics, Medical Faculty RWTH Aachen, University Hospital RWTH Aachen, Aachen, Germany

[*]these authors contributed equally

Corresponding Author:
Carina Benstoem, MSc, PhD
Department of Intensive Care Medicine
Medical Faculty RWTH Aachen
University Hospital RWTH Aachen
Pauwelsstraße 30
Aachen, 52074
Germany
Phone: 49 241 8089471
Email: cbenstoem@ukaachen.de

Abstract

Background: Sepsis is a major health care problem with high morbidity and mortality rates and affects millions of patients. Telemedicine, defined as the exchange of medical information via electronic communication, improves the outcome of patients with sepsis and decreases the mortality rate and length of stay in the intensive care unit (ICU). Additional telemedicine rounds could be an effective component of performance-improvement programs for sepsis, especially in underserved rural areas and hospitals without ready access to critical care physicians.

Objective: Our aim was to evaluate the impact of additional daily telemedicine rounds on adherence to sepsis bundles. We hypothesized that additional telemedicine support may increase adherence to sepsis guidelines and improve the detection rates of sepsis and septic shock.

Methods: We conducted a retrospective, observational, multicenter study between January 2014 and July 2015 with one tele-ICU center and three ICUs in Germany. We implemented telemedicine as part of standard care and collected data continuously during the study. During the daily telemedicine rounds, routine screening for sepsis was conducted and adherence to the Surviving Sepsis Campaign's 3-hour and 6-hour sepsis bundles were evaluated.

Results: In total, 1168 patients were included in this study, of which 196 were positive for severe sepsis and septic shock. We found that additional telemedicine rounds improved adherence to the 3-hour (Quarter 1, 35% vs Quarter 6, 76.2%; *P*=.01) and 6-hour (Quarter 1, 50% vs Quarter 6, 95.2%; *P*=.001) sepsis bundles. In addition, we noted an increase in adherence to the item "Administration of fluids when hypotension" (Quarter 1, 80% vs Quarter 6, 100%; *P*=.049) of the 3-hour bundle and the item "Remeasurement of lactate" (Quarter 1, 65% vs Quarter 6, 100%, *P*=.003) of the 6-hour bundle. The ICU length of stay after diagnosis of severe sepsis and septic shock remained unchanged over the observation period. Due to a higher number of patients with sepsis in Quarter 5 (N=60) than in other quarters, we observed stronger effects of the additional rounds on mortality in this quarter (Quarter 1, 50% vs Quarter 5, 23.33%, *P*=.046).

Conclusions: Additional telemedicine rounds are an effective component of and should be included in performance-improvement programs for sepsis management.

(J Med Internet Res 2019;21(1):e11161) doi:10.2196/11161

KEYWORDS

intensive care; outcome improvement; sepsis; sepsis bundle compliance; SSC; tele-ICU; telemedicine

Introduction

Sepsis is the most-frequent cause of morbidity and mortality in most intensive care units (ICUs) worldwide. The incidence of sepsis is increasing due to the high prevalence of severe comorbidities in the ageing population and the growing bacterial drug resistance [1]. In Germany, the number of cases of sepsis increased by 5.7% annually from 2007 to 2013. Although the mortality rate is constantly decreasing worldwide, it is still high (30%-50%) [2]. A recently published observational study by the SepNet Critical Care Trials Group showed that application of the new Sepsis-3 definition [3] led to higher rates of ICU and in-hospital mortality (>50%) in patients with severe sepsis or septic shock than in those diagnosed using the previous definition [4]. Early detection of sepsis followed by early initiation of adequate management can significantly improve the outcome in patients with sepsis [5].

The European Society of Intensive Care Medicine and the Society of Critical Care Medicine published the Surviving Sepsis Campaign (SSC) guidelines for the management of severe sepsis and septic shock, with an aim to reduce the mortality of sepsis by 25% in 5 years [5,6,7]. A large body of evidence indicates that adherence to clinical practice guidelines and compliance with sepsis bundles are associated with reduced ICU length of stay, low mortality rates, and improved patient outcome [8,9]. In addition, the SSC guidelines specifically emphasize the need for performance-improvement programs for sepsis [5] and recommend an interdisciplinary approach to sepsis management, protocol development, and implementation; evaluation of targeted metrics; continuous data collection; and continuous feedback to allow constant performance improvement. Although the details may vary among different improvement programs, the common goal is to improve compliance with sepsis bundles and clinical practice guidelines [5,6,7]. However, recent studies showed that compliance with sepsis bundles is still low [8,10,11].

Telemedicine, defined as the exchange of medical information via electronic communication, can be used to improve the availability and quality of medical care. In the intensive care setting, telemedicine may improve early detection and appropriate treatment of severe sepsis and septic shock [12]. Telemedicine facilitates direct interaction among intensive care providers as an around-the-clock service over long distances and physicians who care for critically ill patients (intensivist to physician). Telemedicine enables critical decision support by exchanging clinical data in real time [13]. Recently, a systematic review demonstrated that telemedicine can improve the outcome of critically ill patients and decrease ICU mortality and length of stay [14]. However, no study has thus far determined whether additional telemedicine rounds could be an effective component of performance-improvement programs for sepsis, especially in underserved rural areas and hospitals without ready access to critical care physicians. Our objective was to evaluate the impact of additional daily telemedicine rounds via an audio-video system on the adherence to 3-hour and 6-hour sepsis bundles in three ICUs in Germany. We hypothesized that additional telemedicine support will increase adherence to sepsis guidelines and improve the detection rates of severe sepsis and septic shock.

Methods

Study Design and Oversight

We performed an 18-month retrospective, observational, multicenter study in three ICUs within three hospitals in North Rhine Westphalia (Germany). The study started in January 2014 and ended in June 2015 (18 months). We included only adult, nonpregnant patients aged ≥18 years, with no advanced care directives that limited life-saving care in our analysis. The study was reviewed and approved by the local institutional ethics board of the University Hospital RWTH Aachen (262/13). We implemented telemedicine as part of standard of care; as the analysis was performed retrospectively, the ethics board waved the need for informed consent.

Characteristics of Intensive Care Units

One ICU (A), focusing on neurosurgery and general surgery, was located at a University Hospital and staffed by an intensivist around the clock. Two interdisciplinary ICUs (B and C) were located in community hospitals. ICU B was staffed by a general anesthesiologist and an internal specialist on weekdays. After regular day shifts, on-call personnel or anesthesia house staff was responsible for the treatment of ICU patients during the night and on weekends. ICU C was staffed by physicians of the Department of Internal Medicine and by anesthesiologists during regular day shifts; during the night and on weekends, an anesthesiologist was on-call.

Tele-ICU Center and Telemedicine Infrastructure

The Telemedicine Center at the University Hospital RWTH Aachen was the leading center for this study. As a preparatory measure, both the tele-ICU system and the electronic health record (*FallAkte*; Soarian Integrated Care, Siemens, Munich, Germany) were customized for use in the ICUs and the tele-ICU. All participating personnel received standardized information including the SSC guidelines and relevant literature [3,4,5]. Prior to the beginning of the study, a permanently installed tele-ICU system was implemented at the Telemedicine Center in the University Hospital RWTH Aachen and mobile tele-ICU systems were installed at the participating ICUs (Figure 1). A secure encrypted site-to-site virtual private network connection was established. The Telemedicine Center and participating ICUs were equipped with identical audiovisual transmission equipment (Cisco Systems, Inc, San José, CA). In addition, the Telemedicine Center received a workstation, multiple monitors, and a video system (Cisco Systems, Inc). The participating ICUs were equipped with a mobile audio-video system that could be

taken to the patient's room during rounds. Videoconferencing allowed two or more units to communicate simultaneously via a two-way high-resolution video and audio transmission. The video system included an option for closeup zoom that allowed detailed examination of both the patient and the bedside equipment. To enable exchange of medical data, the secure data-protection platform *FallAkte* was established. Accordingly, the system integrated a "store and forward" technique for structured data combined with real-time audio-visual telemedicine rounds, through which the vital signs were measured. After checking for missing data, all patient data were anonymized and exported to the University Hospital's research database. Figure 1 outlines the structure of the Telemedicine Center and the participating ICUs.

The Telemedicine Center was staffed with an intensive care physician from 7:30 AM to 4 PM on weekdays and from 9 AM to 5:30 PM on weekends. Telemedicine rounds were performed on a daily basis. If required, additional rounds were offered after core hours and at night by the intensive care consultant on duty, ensuring accessibility of telemedical consultations around the clock.

Medical documentation in two ICUs was entirely paper based before the study started. To allow standardized documentation, project-specific templates were designed (Adobe Life Cycle, San José, CA) and filled in by physicians in the Telemedicine Center and ICUs during daily rounds. The templates covered basic demographic data of patients and summarized diagnoses, performed procedures, and therapies. Furthermore, the severity of disease and functional limitations of the patient were assessed by the Simplified Acute Physiology Score II (SAPS II) and Sepsis-related Organ Failure Assessment (SOFA) scales [13]. A systematic checklist for daily infection assessment was used. During rounds, both case presentation and assessment were performed by the local physician and the intensivist in the tele-ICU; the diagnostic or therapeutic interventions were discussed, and direct feedback regarding sepsis management was provided. Typically, conscious patients or their relatives were present during the telemedicine rounds.

All telemedicine rounds were observed and documented by a research assistant. The following items were documented: recommendations for diagnosis and therapy of sepsis, adjustments of sepsis management, details regarding antibiotic therapy (continued, evaluated, changed, or terminated), scoring of SAPS II and SOFA, duration of ICU length of stay (LOS), and duration of ICU LOS after sepsis diagnosis.

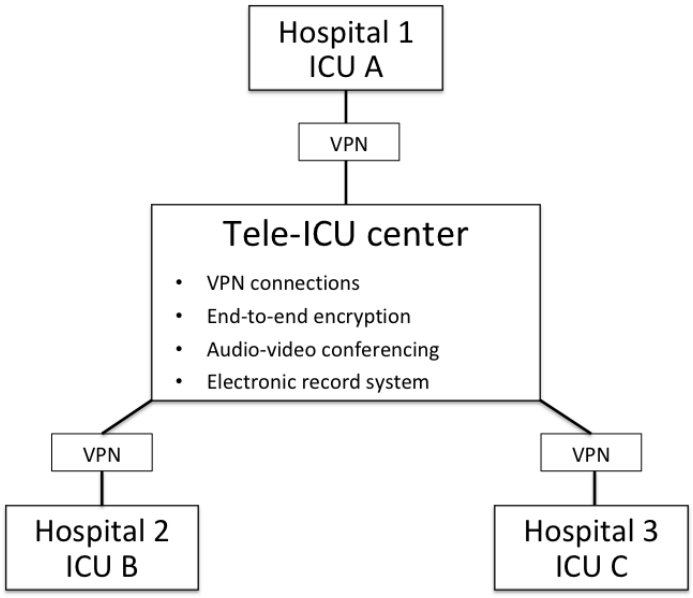
Definitions

The applied definitions of sepsis, severe sepsis, and septic shock are outlined in Multimedia Appendix 1 [7,15].

Evaluation of Adherence to 3-Hour and 6-Hour Sepsis Bundles in Patients With Sepsis

We continuously extracted data during the study and follow-up to evaluate whether sepsis management fulfilled the requirements of the 3-hour and 6-hour sepsis bundles. Time "0" was defined as the time when the attending physician diagnosed sepsis. Items of the 3-hour and 6-hour bundles for patients with severe sepsis and septic shock were adapted from the SSC standard sepsis resuscitation guidelines. The 3-hour bundle includes the following recommendations: measurement of lactate levels, blood cultures obtained prior to administration of antibiotics, administration of broad spectrum antibiotics, and administration of 20 mL/kg crystalloid fluid for hypertension or lactate levels ≥4 mmol/L. The 6-hour bundle consisted of the following core items: application of vasopressors for persistent hypotension (mean arterial pressure <65 mmHg) despite initial fluid resuscitation, assessment of central venous pressure and central venous oxygen saturation if hypotension persisted despite initial fluid administration or when initial lactate levels were ≥4 mmol/L, and remeasurement of lactate levels at initial elevation. The second bundle was also applied to patients with persistent hypotension or high lactate levels within the 6-hour period.

Figure 1. Outline of the telemedicine center and the participating ICUs. ICU: intensive care unit, VPN: virtual private network.



For evaluation of adherence to the 3-hour and 6-hour sepsis bundles, we assigned a "yes" rating to patients with sepsis if all core items of the respective bundle were executed within 3-hours or 6-hours after time "0"; otherwise, a "no" rating was assigned.

Statistical Analysis

Categorical data are presented as frequency and percentage. Frequencies of categorical data were compared between groups by the Fisher exact test. Continuous variables are expressed as mean values (SD). Differences in continuous data between groups were analyzed by the *t* test, assuming unequal variances. Statistical tests were two-tailed, and values of $P<.05$ were considered significant. SAS software, version 9.4 (SAS Institute Inc, Cary, NC), and GraphPad Prism software, version 6.0 (GraphPad Software, La Jolla, CA), were used for statistical analyses.

Results

The study included 1168 patients who received 4569 telemedicine rounds in addition to their daily rounds at the local ICUs. Physicians at the 3 ICUs performed 4373 infections and sepsis screenings. The telemedicine rounds were performed routinely in the morning when it was most convenient for the staff of the participating ICU. There were rarely emergency visits outside the core visit times. Emergency telemedicine visits were infrequent, especially for patients with acute respiratory distress syndrome, craniocerebral trauma, or septic shock, and the involved staff scheduled these events, if needed. During the study, the average visiting time was 4.67 (SD 2.55) minutes. The duration of the telemedicine rounds was comparable among patients during the course of the study (range of mean: 4.18-4.63) with similar variation (range of SD: 2.27-2.82).

Overall, we observed a decrease in mortality from 50% in Quarter 1 (January 1, 2014 to March 31, 2014) to 33.33% in Q6 (April 1, 2015 to June 30, 2015; $P=.35$) in patients with severe sepsis and septic shock and a comparable degree of severity among patients, as per the SAPS II and SOFA score (Tables 1 and 2). Due to a higher number of patients with sepsis in Quarter 5 (N=60) than in other quarters, we observed stronger and statistically significant effects of the additional rounds on mortality in this quarter (Quarter 1, 50% vs Quarter 5, 23.33%, $P=.046$). ICU LOS after diagnosis of severe sepsis and septic shock remained unchanged over the observation period. We presented the number of scorings, number of therapeutic recommendations made, and the overall number of sepsis detections per quarter in Table 1. The mean age of the included patients was 64.91 (SD 17.09) years, and 584 of 1168 patients (50%) were male. In total, 196 patients showed positive results for severe sepsis (N=95) or septic shock (N=101) during the study period and were included in our analysis. Table 2 summarizes the patient characteristics and provides details on ICU LOS and ICU LOS after diagnosis of sepsis for Quarters 1, 5, and 6.

Our primary objective was to evaluate the impact of additional daily telemedicine rounds on adherence to the sepsis bundles in order to determine whether additional telemedicine rounds are an effective performance-improvement strategy for sepsis management. We found that additional telemedicine rounds had a statistically significant effect on adherence to the 3-hour ($P=.01$) and the 6-hour ($P=.001$) sepsis bundles. In addition, we observed an increase in adherence to the item "Administration of fluids when hypotension" ($P=.049$) with the 3-hour bundle and to the item "Remeasurement of lactate" ($P=.003$) with the 6-hour bundle. All results of the impact of telemedicine rounds on adherence to sepsis bundles are presented in Table 3 and Figure 2. Figure 3 illustrates the change in mortality among patients diagnosed with sepsis.

Table 1. Study characteristics.

Study characteristics	Quarter 1, n (%)	Quarter 2, n (%)	Quarter 3, n (%)	Quarter 4, n (%)	Quarter 5, n (%)	Quarter 6, n (%)	Total (N)
Telemedicine rounds	541 (11.84)	596 (13.04)	953 (20.86)	812 (17.77)	1074 (23.50)	593 (12.98)	4569
Infection and sepsis screenings	424 (9.70)	591 (13.51)	990 (22.64)	775 (17.72)	1030 (23.55)	563 (12.87)	4373
Scorings (SAPS[a] II and SOFA[b])	541 (11.84)	596 (13.04)	953 (20.86)	812 (17.77)	1074 (23.51)	593 (12.98)	4569
Diagnostic recommendations	90 (12.21)	77 (10.45)	162 (21.98)	108 (14.65)	203 (27.54)	97 (13.16)	737
Therapeutic recommendations	111 (9.28)	82 (6.85)	285 (23.82)	202 (16.89)	363 (30.35)	153 (12.79)	1196
Total 3-h and 6-h bundles	20 (10.20)	18 (9.18)	36 (18.37)	41 (20.92)	60 (30.61)	21 (10.71)	196
Severe sepsis or septic shock detections	20 (10.20)	18 (9.18)	36 (18.37)	41 (20.92)	60 (30.61)	21 (10.71)	196
Mortality	10/20 (50)	8/18 (44.44)	14/36 (38.88)	10/41 (24.39)	14/60 (23.33)	7/21 (33.33)	63

[a]SAPS: Simplified Acute Physiology Score.

[b]SOFA: Sepsis-related Organ Failure Assessment.

Table 2. Characteristics of patients with severe sepsis and septic shock.

Patient characteristics	Initiation of telemedicine rounds, Quarter 1 (n=20)	After implementation in Quarter 5 (n=60)	After implementation in Quarter 6 (n=21)	*P* value	
				Comparison between Q1 and Q5	Comparison between Q1 and Q6
Patients with severe sepsis, n (%)	10 (50.0)	34 (56.7)[a]	6 (28.6)[b]	.62	.21
Patients with septic shock, n (%)	10 (50.0)	26 (43.3)[a]	15 (71.4)[b]	.62	.21
Mortality, n (%)	10 (50.0)	14 (23.3)[a]	7 (33.3)[b]	. 046	.35
LOS[c] ICU[d] (days), mean (SD)	18.2 (21.6)	15.65 (15.5)[a]	19.48 (21.4)[b]	.63	.85
LOS ICU after diagnosis of sepsis (days), mean (SD)	15.65 (21.1)	13.22 (13.9)[a]	16.76 (20.7)[b]	.64	.87
SAPS[e] II, mean (SD)	44.35 (12.1)	44.16 (16.4)[a]	45.76 (14.4)[b]	.96	.74
SOFA[f], mean (SD)	7.7 (3.1)	7.18 (3.9)[a]	7.52 (3.6)[b]	.55	.86

[a]Comparison between Q1 and Q5.

[b]Comparison between Q1 and Q6.

[c]LOS: length of stay.

[d]ICU: intensive care unit.

[e]SAPS: Simplified Acute Physiology Score.

[f]SOFA: Sepsis-related Organ Failure Assessment.

Table 3. Impact of telemedicine rounds on adherence to sepsis bundles.

Parameters	Quarter 1[a] (N=20), n (%)	Quarter 6[b] (N=21), n (%)	*P* value
Compliance to the 3-h bundle	7 (35.0)	16 (76.2)	.01
Compliance to the 6-h bundle	10 (50.0)	20 (95.2)	.001
Components or target values of the 3-h bundle			
Serum lactate measurement	20 (100.0)	21 (100.0)	>.99
Blood cultures before antibiotics	11 (55.0)	16 (76.2)	.20
Administration of antibiotics within the first 3 h	19 (95.0)	21 (100.0)	.49
Administration of fluids during hypotension	16 (80.0)	21 (100.0)	.049
Administration of vasopressors when indicated	18 (90.0)	20 (95.2)	.61
CVP[c] >8 mmHg	16 (80.0)	19 (90.5)	.34
$ScvO_2$[d] >70%	5 (25.0)	9 (42.9)	.33
Components or target values of the 6-h bundle			
Administration of vasopressors when indicated	18 (90.0)	20 (95.2)	.61
Assessment of CVP when indicated	16 (80.0)	19 (90.5)	.41
Assessment of $ScvO_2$ when indicated	4 (20.0)	9 (42.9)	.18
Remeasurement of lactate	13 (65.0)	21 (100.0)	.003

[a]Initiation of telemedicine rounds.

[b]After implementation of additional rounds.

[c]CVP: central venous pressure.

[d]$ScvO_2$: central venous oxygen saturation.

Figure 2. Impact of additional telemedicine rounds on adherence to sepsis bundles.

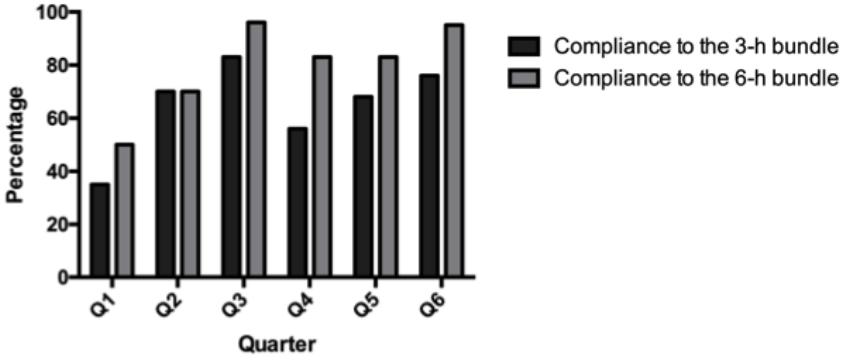
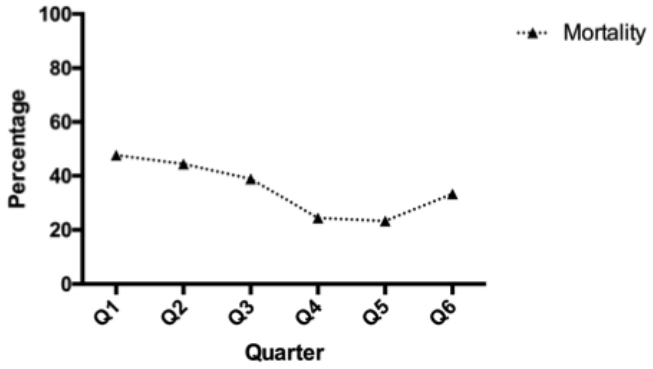


Figure 3. Impact of additional telemedicine rounds on mortality.



Discussion

To the best of our knowledge, this is the first study to evaluate whether additional telemedicine rounds should be included in performance-improvement strategies for sepsis management, especially in underserved rural areas and hospitals without ready 24-hour access to critical care physicians. Of the 1168 patients included in this study, 196 were positive for severe sepsis and septic shock. The proportion of patients with severe sepsis and septic shock was in the expected range and in line with the findings of recent epidemiologic surveys [16,17,18]. We found that additional telemedicine rounds had a significant effect on the adherence to the 3-hour and the 6-hour sepsis bundles and improved adherence to current clinical practice guidelines and patient care in sepsis management. In addition, we observed an increase in adherence to the item “Administration of fluids when hypotension” of the 3-hour bundle and the item “Re-measurement of lactate” in the 6-hour bundle. Moreover, we observed a decrease in mortality among patients with severe sepsis and septic shock and a comparable degree of severity among all patients, as per the SAPS II and SOFA scores. The ICU LOS after diagnosis of severe sepsis and septic shock remained unchanged over the observation period. Due to a higher number of patients with sepsis in Quarter 5 (N=60) than in other quarters, we observed stronger effects of the additional rounds on mortality in this quarter. Our findings are consistent with those of other studies or health care approaches investigating sepsis management in the intensive care medical setting [19,20,21].

Additional telemedicine rounds with standardized daily sepsis screening significantly improved guideline adherence. Near real-time feedback in an intensivist-driven tele-ICU system is an effective performance-improvement strategy for rapid implementation of evidence-based practice to achieve improved quality of care. In Germany, the foundation of telemedicine was laid when the first guideline on telemedicine in intensive care was published by the Association of the Scientific Medical Societies (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften) and the German Society of Anaesthesiology and Intensive Care Medicine (Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin) [22]. Previous studies showed that performance-improvement programs are associated with increased adherence to resuscitation and management of sepsis bundles, with reduced mortality in patients with sepsis, severe sepsis, or septic shock [23]. As a continuous effort, all involved stakeholders should aim to implement telemedicine nationwide as a part of the daily routine standard of care. Constant measurement of the influence of telemedicine on quality indicators for intensive care is important, as demonstrated in this study.

The implementation of telemedicine is important in light of recent findings of Faine and colleagues [24], who showed that interhospital transfers delay appropriate treatment for patients with severe sepsis and septic shock. Telemedicine can improve on-site quality of care by shared expertise, as it is a viable alternative to urgent patient transfers to university centers. As recently demonstrated by Pannu et al [25], telemedicine comanagement slightly increased patient transfers to high-level centers. However, our findings are consistent with those reported

by two systematic reviews and meta-analyses and multiple study reports, which revealed that telemedicine is associated with low ICU mortality [26,27,28,29] and low ICU LOS [26,29,30]. Furthermore, Lilly et al emphasized that telemedicine has the potential to improve adherence to ICU best practices, reduce response times to alarms, and encourage the use of performance data in the ICU [27].

Owing to its explorative nature, this study had a few limitations that should be considered when interpreting our results. First, we only included a small number of participating hospitals, resulting in a small numbers of patients. However, these are preliminary results and provide crucial data for subsequent large-scale trials. Second, the level of acceptance, an important factor for the success for telemedicine, especially in the ICU, was not measured systematically in this study. Therefore, we could not estimate the possible influence of telemedicine as a cofounding factor. However, personal communication during the telemedicine rounds revealed positive feedback and a high level of acceptance by the participating physicians, nurses, patients, and their relatives. Our experience is in line with the findings of other telemedicine studies [30,31] that reported high levels of staff acceptance of telemedicine and tele-ICU coverage. Similar to other medical fields [32], the overall attitude toward telemedicine and eHealth in the intensive care setting was consistently positive and in favor of the technology. This aspect should be assessed as a possible confounding factor in rigorously planned, methodological, high-quality studies in the future. Third, the geographical and time-specific design may limit the extrapolation of our results to other medical centers and patients. We believe that an advanced technical setup will further improve the acceptance of telemedicine by, for example, reducing the workload of documentation. Automatic data capture by export of the international Health Level 7 standard into an active tele-ICU system with automatic calculation of disease severity or sepsis alerts may improve care beyond the results of this study. Fourth, it is important to carefully consider the generalizability of data obtained from a retrospective study such as this study. However, we continuously documented sepsis onset and sepsis bundle compliance throughout our study. This approach reduced difficulties associated with the retrospective identification of the time of sepsis onset as “time zero” for the evaluation of sepsis bundle compliance. This approach might limit the explanatory power of effect size and causality of the tele-ICU concept, but offers a relevant benefit of positive outcomes for patients enrolled in such an implementation study. The last limitation was with regard to some methodological characteristics of our study design. Our retrospective data analysis was based on the evaluation of six consecutive quarters after implementing additional telemedicine rounds as a part of standard care for ICU patients. Patients were included continuously over the course of the six quarters and the number of rounds and total number of infection and sepsis screenings were recorded per quarter. However, in one ICU, some patients received additional rounds in two subsequent quarters. As such, we could not present some data items specific to one quarter, which limited our ability to perform a pre-post comparison or patient-level logistic regression analysis including adjustment for patient demographics, comorbidities, or sepsis severity. However, we believe that our findings, which showed a significant advance in sepsis management over the course of six quarters with telemedicine support, supports the implementation of additional telemedicine rounds as a successful performance-improvement strategy for sepsis management.

Acknowledgments

This study was funded by the Ministry of Health, Emancipation, Care and the Elderly of North Rhine Westphalia under the “ZIEL-2-Programm” and cofunded by the European Regional Development Fund.

Conflicts of Interest

GM received research grants, travel reimbursements, and consultancy fees from B Braun Melsungen (Niedersachsen, Germany), Biotest (Hessen, Germany), and Adrenomed AG (Brandenburg, Germany). The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Multimedia Appendix 1

Definitions of sepsis, severe sepsis or septic shock.

[PDF File (Adobe PDF File), 31KB - jmir_v21i1e11161_app1.pdf]

References

1. Kempker JA, Martin GS. The Changing Epidemiology and Definitions of Sepsis. Clin Chest Med 2016 Dec;37(2):165-179 [FREE Full text] [doi: 10.1016/j.ccm.2016.01.002] [Medline: 27229635]
2. Fleischmann C, Thomas-Rueddel DO, Hartmann M, Hartog CS, Welte T, Heublein S, et al. Hospital Incidence and Mortality Rates of Sepsis. Dtsch Arztebl Int 2016 Mar 11;113(10):159-166 [FREE Full text] [doi: 10.3238/arztebl.2016.0159] [Medline: 27010950]
3. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016 Feb 23;315(8):801-810 [FREE Full text] [doi: 10.1001/jama.2016.0287] [Medline: 26903338]

4. SepNet Critical Care Trials Group. Incidence of severe sepsis and septic shock in German intensive care units: the prospective, multicentre INSEP study. Intensive Care Med 2016 Dec;42(12):1980-1989. [doi: 10.1007/s00134-016-4504-3] [Medline: 27686355]
5. Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, Cohen J, et al. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. Intensive Care Med 2004 Apr;30(4):536-555. [doi: 10.1007/s00134-004-2210-z] [Medline: 14997291]
6. Levy MM, Pronovost PJ, Dellinger RP, Townsend S, Resar RK, Clemmer TP, et al. Sepsis change bundles: converting guidelines into meaningful change in behavior and clinical outcome. Crit Care Med 2004 Nov;32(11 Suppl):S595-S597. [Medline: 15542969]
7. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Surviving Sepsis Campaign Guidelines Committee including The Pediatric Subgroup. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012. Intensive Care Med 2013 Feb;39(2):165-228. [doi: 10.1007/s00134-012-2769-8] [Medline: 23361625]
8. Rhodes A, Phillips G, Beale R, Cecconi M, Chiche JD, De Backer D, et al. The Surviving Sepsis Campaign bundles and outcome: results from the International Multicentre Prevalence Study on Sepsis (the IMPreSS study). Intensive Care Med 2015 Sep;41(9):1620-1628. [doi: 10.1007/s00134-015-3906-y] [Medline: 26109396]
9. Levy MM, Rhodes A, Phillips GS, Townsend SR, Schorr CA, Beale R, et al. Surviving Sepsis Campaign: association between performance metrics and outcomes in a 7.5-year study. Intensive Care Med 2014 Nov;40(11):1623-1633. [doi: 10.1007/s00134-014-3496-0] [Medline: 25270221]
10. Brunkhorst FM, Engel C, Ragaller M, Welte T, Rossaint R, Gerlach H, German Sepsis Competence Network (SepNet). Practice and perception--a nationwide survey of therapy habits in sepsis. Crit Care Med 2008 Oct;36(10):2719-2725. [doi: 10.1097/CCM.0b013e318186b6f3] [Medline: 18766100]
11. Bloos F, Thomas-Rüddel D, Rüddel H, Engel C, Schwarzkopf D, Marshall JC, MEDUSA Study Group. Impact of compliance with infection management guidelines on outcome in patients with severe sepsis: a prospective observational multi-center study. Crit Care 2014 Mar 03;18(2):R42 [FREE Full text] [doi: 10.1186/cc13755] [Medline: 24589043]
12. Cummings J, Krsek C, Vermoch K, Matuszewski K, University HealthSystem Consortium ICU Telemedicine Task Force. Intensive care unit telemedicine: review and consensus recommendations. Am J Med Qual 2007;22(4):239-250. [doi: 10.1177/1062860607302777] [Medline: 17656728]
13. Nguyen Y, Kahn JM, Angus DC. Reorganizing adult critical care delivery: the role of regionalization, telemedicine, and community outreach. Am J Respir Crit Care Med 2010 Jun 01;181(11):1164-1169 [FREE Full text] [doi: 10.1164/rccm.200909-1441CP] [Medline: 20224067]
14. Wilcox ME, Adhikari NKJ. The effect of telemedicine in critically ill patients: systematic review and meta-analysis. Crit Care 2012 Jul 18;16(4):R127 [FREE Full text] [doi: 10.1186/cc11429] [Medline: 22809335]
15. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, SCCM/ESICM/ACCP/ATS/SIS. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. Crit Care Med 2003 Apr;31(4):1250-1256. [doi: 10.1097/01.CCM.0000050454.01978.3B] [Medline: 12682500]
16. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996 Jul;22(7):707-710. [Medline: 8844239]
17. Engel C, Brunkhorst FM, Bone H, Brunkhorst R, Gerlach H, Grond S, et al. Epidemiology of sepsis in Germany: results from a national prospective multicenter study. Intensive Care Med 2007 Apr;33(4):606-618. [doi: 10.1007/s00134-006-0517-7] [Medline: 17323051]
18. SepNet Critical Care Trials Group. Incidence of severe sepsis and septic shock in German intensive care units: the prospective, multicentre INSEP study. Intensive Care Med 2016 Dec;42(12):1980-1989. [doi: 10.1007/s00134-016-4504-3] [Medline: 27686355]
19. Loyola S, Wilhelm J, Fornos J. An innovative approach to meeting early goal-directed therapy using telemedicine. Crit Care Nurs Q 2011;34(3):187-199. [doi: 10.1097/CNQ.0b013e31821df105] [Medline: 21670617]
20. Badawi O, Hassan E. Telemedicine and the patient with sepsis. Crit Care Clin 2015 Apr;31(2):291-304. [doi: 10.1016/j.ccc.2014.12.007] [Medline: 25814455]
21. Machado SM, Wilson EH, Elliott JO, Jordan K. Impact of a telemedicine eICU cart on sepsis management in a community hospital emergency department. J Telemed Telecare 2018 Apr;24(3):202-208. [doi: 10.1177/1357633X17691862] [Medline: 29278979]
22. Marx G, Koch T. Telemedizin in der Intensivmedizin Strukturempfehlungen der DGAI*. Anästh Intensivmed 2015;56:257-261 [FREE Full text]
23. Damiani E, Donati A, Serafini G, Rinaldi L, Adrario E, Pelaia P, et al. Effect of performance improvement programs on compliance with sepsis bundles and mortality: a systematic review and meta-analysis of observational studies. PLoS One 2015;10(5):e0125827 [FREE Full text] [doi: 10.1371/journal.pone.0125827] [Medline: 25946168]
24. Faine BA, Noack JM, Wong T, Messerly JT, Ahmed A, Fuller BM, et al. Interhospital Transfer Delays Appropriate Treatment for Patients With Severe Sepsis and Septic Shock: A Retrospective Cohort Study. Crit Care Med 2015 Dec;43(12):2589-2596. [doi: 10.1097/CCM.0000000000001301] [Medline: 26491865]

25. Pannu J, Sanghavi D, Sheley T, Schroeder DR, Kashyap R, Marquez A, et al. Impact of Telemedicine Monitoring of Community ICUs on Interhospital Transfers. Crit Care Med 2017 Aug;45(8):1344-1351 [FREE Full text] [doi: 10.1097/CCM.0000000000002487] [Medline: 28481753]
26. Young LB, Chan PS, Lu X, Nallamothu BK, Sasson C, Cram PM. Impact of telemedicine intensive care unit coverage on patient outcomes: a systematic review and meta-analysis. Arch Intern Med 2011 Mar 28;171(6):498-506. [doi: 10.1001/archinternmed.2011.61] [Medline: 21444842]
27. Lilly CM, McLaughlin JM, Zhao H, Baker SP, Cody S, Irwin RS, UMass Memorial Critical Care Operations Group. A multicenter study of ICU telemedicine reengineering of adult critical care. Chest 2014 Mar 01;145(3):500-507. [doi: 10.1378/chest.13-1973] [Medline: 24306581]
28. Wilcox ME, Adhikari NKJ. The effect of telemedicine in critically ill patients: systematic review and meta-analysis. Crit Care 2012 Jul 18;16(4):R127 [FREE Full text] [doi: 10.1186/cc11429] [Medline: 22809335]
29. Sadaka F, Palagiri A, Trottier S, Deibert W, Gudmestad D, Sommer SE, et al. Telemedicine intervention improves ICU outcomes. Crit Care Res Pract 2013;2013:456389 [FREE Full text] [doi: 10.1155/2013/456389] [Medline: 23365729]
30. Young LB, Chan PS, Cram P. Staff acceptance of tele-ICU coverage: a systematic review. Chest 2011 Feb;139(2):279-288 [FREE Full text] [doi: 10.1378/chest.10-1795] [Medline: 21051386]
31. Müller KI, Alstadhaug KB, Bekkelund SI. Acceptability, Feasibility, and Cost of Telemedicine for Nonacute Headaches: A Randomized Study Comparing Video and Traditional Consultations. J Med Internet Res 2016 May 30;18(5):e140 [FREE Full text] [doi: 10.2196/jmir.5221] [Medline: 27241876]
32. Kessel KA, Vogel MME, Schmidt-Graf F, Combs SE. Mobile Apps in Oncology: A Survey on Health Care Professionals' Attitude Toward Telemedicine, mHealth, and Oncological Apps. J Med Internet Res 2016 Dec 24;18(11):e312 [FREE Full text] [doi: 10.2196/jmir.6399] [Medline: 27884810]

Abbreviations

CVP: central venous pressure
ICU: intensive care unit
LOS: length of stay
$ScvO_2$: central venous oxygen saturation
SOFA: Sepsis-related Organ Failure Assessment
SAPS: Simplified Acute Physiology Score
SSC: Surviving Sepsis Campaign

Edited by G Eysenbach; submitted 29.05.18; peer-reviewed by S Tafelski, S Manaktala; comments to author 19.07.18; revised version received 20.08.18; accepted 08.10.18; published 15.01.19.

Please cite as:
Deisz R, Rademacher S, Gilger K, Jegen R, Sauerzapfe B, Fitzner C, Stoppe C, Benstoem C, Marx G
Additional Telemedicine Rounds as a Successful Performance-Improvement Strategy for Sepsis Management: Observational Multicenter Study
J Med Internet Res 2019;21(1):e11161
URL: http://www.jmir.org/2019/1/e11161/
doi:10.2196/11161
PMID:30664476

Original Paper

Use of the CHA2DS2-VASc Score for Risk Stratification of Hospital Admissions Among Patients With Cardiovascular Diseases Receiving a Fourth-Generation Synchronous Telehealth Program: Retrospective Cohort Study

Jen-Kuang Lee[1,2,3], MD, PhD; Chi-Sheng Hung[1,2], MD, PhD; Ching-Chang Huang[1,2], MD; Ying-Hsien Chen[1,2], MD; Pao-Yu Chuang[1,4], MSN; Jiun-Yu Yu[1,5], PhD; Yi-Lwun Ho[1,2], MD, PhD

[1]Telehealth Center, National Taiwan University Hospital, Taipei, Taiwan

[2]Division of Cardiology, Department of Internal Medicine, National Taiwan University College of Medicine and Hospital, Taipei, Taiwan

[3]Department of Laboratory Medicine, National Taiwan University College of Medicine and Hospital, Taipei, Taiwan

[4]Department of Nursing, National Taiwan University Hospital, Taipei, Taiwan

[5]Department of Business Administration, College of Management, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Yi-Lwun Ho, MD, PhD
Telehealth Center
National Taiwan University Hospital
No 7, Chung-Shan South Road
Taipei,
Taiwan
Phone: 886 2 23123456 ext 63651
Fax: 886 2 27044688
Email: ylho@ntu.edu.tw

Abstract

Background: Telehealth programs are generally diverse in approaching patients, from traditional telephone calling and texting message and to the latest fourth-generation synchronous program. The predefined outcomes are also different, including hypertension control, lipid lowering, cardiovascular outcomes, and mortality. In previous studies, the telehealth program showed both positive and negative results, providing mixed and confusing clinical outcomes. A comprehensive and integrated approach is needed to determine which patients benefit from the program in order to improve clinical outcomes.

Objective: The CHA_2DS_2-VASc (congestive heart failure, hypertension, age >75 years [doubled], type 2 diabetes mellitus, previous stroke, transient ischemic attack or thromboembolism [doubled], vascular disease, age of 65-75 years, and sex) score has been widely used for the prediction of stroke in patients with atrial fibrillation. This study investigated the CHA_2DS_2-VASc score to stratify patients with cardiovascular diseases receiving a fourth-generation synchronous telehealth program.

Methods: This was a retrospective cohort study. We recruited patients with cardiovascular disease who received the fourth-generation synchronous telehealth program at the National Taiwan University Hospital between October 2012 and June 2015. We enrolled 431 patients who had joined a telehealth program and compared them to 1549 control patients. Risk of cardiovascular hospitalization was estimated with Kaplan-Meier curves. The CHA_2DS_2-VASc score was used as the composite parameter to stratify the severity of patients' conditions. The association between baseline characteristics and clinical outcomes was assessed via the Cox proportional hazard model.

Results: The mean follow-up duration was 886.1 (SD 531.0) days in patients receiving the fourth-generation synchronous telehealth program and 707.1 (SD 431.4) days in the control group (P<.001). The telehealth group had more comorbidities at baseline than the control group. Higher CHA_2DS_2-VASc scores (≥4) were associated with a lower estimated rate of remaining free from cardiovascular hospitalization (46.5% vs 54.8%, log-rank P=.003). Patients with CHA_2DS_2-VASc scores ≥4 receiving the telehealth program were less likely to be admitted for cardiovascular disease than patients not receiving the program. (61.5% vs 41.8%, log-rank P=.01). The telehealth program remained a significant prognostic factor after multivariable Cox analysis in patients with CHA_2DS_2-VASc scores ≥4 (hazard ratio=0.36 [CI 0.22-0.62], P<.001)

Conclusions: A higher CHA_2DS_2-VASc score was associated with a higher risk of cardiovascular admissions. Patients accepting the fourth-generation telehealth program with CHA_2DS_2-VASc scores ≥4 benefit most by remaining free from cardiovascular hospitalization.

(J Med Internet Res 2019;21(1):e12790) doi:10.2196/12790

KEYWORDS

CHA2DS2-VASc score; fourth-generation synchronous telehealth program; hospitalization; cardiovascular disease

Introduction

Cardiovascular diseases remain the biggest health burden worldwide [1]. Telemedicine can be used to monitor disease and treat patients in real-time. Previous studies showed that patients with cardiovascular disease accepting telehealth medicine had better control of vascular risk factors such as hypertension, diabetes mellitus, and dyslipidemia [2]. Both Spyros et al and Sally et al reported that telemedicine was an important prognostic factor for reducing all-cause mortality in patients with congestive heart failure [3,4]. We have shown that the fourth-generation telehealth program—an internet-based, synchronized, disease-management program providing an immediate response—could lower mortality as compared to the control group [5]. However, Takahashi et al found that telemonitoring did not lead to fewer hospitalizations or emergency department visits [6]. A review article suggested that telemedicine should be carefully evaluated and applied to patients who benefit with improved clinical outcomes [7].

The European Society of Cardiology has published atrial fibrillation-management guidelines in 2016 to advocate using the CHA_2DS_2-VASc (congestive heart failure, hypertension, age >75 years [doubled], type 2 diabetes mellitus, previous stroke, transient ischemic attack or thromboembolism [doubled], vascular disease, age of 65-75 years, and sex) score as a predictive scoring model for stroke in patients with atrial fibrillation [8]. In addition, the American Heart Association has already released similar recommendations for treatment of patients with atrial fibrillation in 2014 [9]. Based on the guidelines, patients with one or more stroke risk factors (ie, a CHA_2DS_2-VASc score ≥1 in men or ≥2 in women) are at a higher risk for future stroke events, which is a dosage effect. Oral anticoagulation is recommended or preferred for these patients with atrial fibrillation. Furthermore, Mitchell et al showed that CHA_2DS_2-VASc scores could predict the incidence of stroke or transient ischemic attack in a population of over 20,000 patients with acute coronary syndrome without atrial fibrillation [10]. This scoring system can also predict various categories of cardiovascular hospitalization events other than cerebrovascular accident events; the scores themselves were a composite of inflammatory risk factors [11]. However, no studies have thus far addressed the utility of CHA_2DS_2-VASc scores to stratify patients via a telehealth program.

Here, we aimed to investigate the relation between CHA_2DS_2-VASc scores and cardiovascular admission in patients receiving a fourth-generation telehealth program. In addition, we stratified patients by CHA_2DS_2-VASc score and examined the effect of the telehealth program on these scores and clinical outcomes.

Methods

Study Design

This was a single-center, clinical retrospective epidemiologic study that was approved by the Institutional Review Board of National Taiwan University Hospital, Taipei, Taiwan. All clinical managements of patients in the telehealth program were performed in accordance with relevant guidelines and regulations.

Patient Selection

The study was conducted from October 2012 to June 2015 at the Telehealth Center of the hospital by the Taiwan ELEctroHEALTH Study Group (TELEHEALTH Study Group). Patients older than 20 years diagnosed with chronic cardiovascular diseases and receiving the telehealth program at our telehealth center were enrolled as the study group. The telehealth-care program is a self-pay service in our hospital, which is not reimbursed by health insurance. Because the patients needed to pay for the service and receive long-term follow-up, we only included patients who were above 20 years old in this study. The decision of receiving the telehealth program depended on the patients or their caregivers. Chronic cardiovascular diseases included coronary artery disease, myocardial infarction, heart failure, peripheral artery disease, stroke, and hypertension. The control group included participants who visited our cardiovascular center during the same period but did not participate in the telehealth care program (received usual care only). The exclusion criteria in this study (for both telehealth group and control group) were age <20 years, absence of any one of the abovementioned chronic cardiovascular diseases, and no follow-up in our hospital.

Telehealth Care Program

The fourth-generation telehealth program at our center is a synchronized and integrated remote management program for chronic diseases. The internet-based platform was developed by the Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taiwan. The details of this program have been reported previously [12]. Briefly, this telehealth program provides the following services: biometric data including single-lead electrocardiography, blood pressure, heart rate, and oximetry are transferred from patients to our telehealth center daily and on demand; nurse case managers telephone patients daily on demand for communication and health promotion; full-time nurse case managers and cardiologists are in charge of care 24 hours per day; and

long-term medication and management are discussed with the patients' primary care physician after acute events. This telehealth program bridges acute and home care and emphasizes on education, prevention, and early detection of clinical deterioration. The clinical information including CHA_2DS_2-VASc scores were relayed to the cardiology specialist who made the final judgment and suggestions regarding care.

Usual Care

Patients in the control group received the usual care provided by the primary care physicians at our cardiovascular center according to updated guidelines including, but not limited to, the American Heart Association's guidelines for lifestyle modification and primary prevention to reduce cardiovascular risk, guidelines for the management of stable ischemic heart disease, and the American Diabetes Association's guidelines for the management of diabetes. Patients made routine outpatient department visits (once every 3 months) to their primary care physicians. There was no contact between the telehealth center and patients receiving usual care.

Data Collection

All demographic and clinical data were obtained from the electronic database of the hospital. The CHA_2DS_2-VASc score was calculated retrospectively according to documentation of the electrical medical chart for congestive heart failure, hypertension, age >75 years (doubled), type 2 diabetes mellitus, previous stroke, transient ischemic attack or thromboembolism (doubled), vascular disease, age of 65-75 years, and sex category. The calculation is not yet an automated process in our Web-based telehealth program. The diagnosis of a chronic disease was based on the electronic database. The discharge diagnosis was used if there was disagreement between outpatient and discharge diagnoses. The follow-up data were acquired from the electronic database of our hospital. The primary outcome of this study was hospitalization for cardiovascular events including acute coronary syndrome, peripheral artery disease, stroke, transient ischemic attack, congestive heart failure, atrial fibrillation, and sudden cardiac death. The end date of follow-up was the September 30, 2016.

Statistical Analysis

Normally distributed data were displayed as the mean (SD), and data were compared within the study group using the *t* test of variance. Nonnormally distributed continuous data were displayed as the median (interquartile range) and were compared in the study group using the Kruskal-Wallis analysis. The distribution of categorical variables was compared in the study group using the chi-squared test. There were no missing values in the basic variable collections in our cohort.

Kaplan-Meier curves were used to estimate survival rates of hospitalization for cardiovascular events, and a log-rank test was used to compare risks among the study groups. We further stratified the subjects according to the CHA_2DS_2-VASc scores of 0-3 and 4-8. To evaluate the independent effect of the telehealth program on the risks of hospitalization for cardiovascular events, we used a multivariable Cox proportional hazard model with adjustment for prespecified clinical characteristics including age >80 years, female gender, diabetes mellitus, hypertension, dyslipidemia, coronary artery disease, congestive heart failure, peripheral artery disease, ischemic stroke, atrial fibrillation, chronic kidney disease, and telehealth program. The assumption of proportional hazard was tested by the Schoenfeld partial residuals, in which the study group was the only explanatory continuous variable. The assumption of proportional hazard was not rejected. The Bonferroni correction was used to adjust for multiple (pairwise) comparisons in the study group when the overall test was statistically significant. Data were analyzed using Statistical Package for Social Science (version 22; IBM Corp, Armonk, NY). Statistical significance was set at two-sided *P* values <.05.

Results

Patient Demographics and Clinical Features

A total of 1980 patients (431 in the telehealth group and 1549 in the control group) were enrolled in this study (Figure 1). The baseline characteristics are reported in the Table 1. In the telehealth groups, the mean age was 70.3 (SD 14.9) years, and 66.1% (285/431) were men.

Figure 1. Flow chart of enrolled patients. CV: cardiovascular; f/u: follow up.

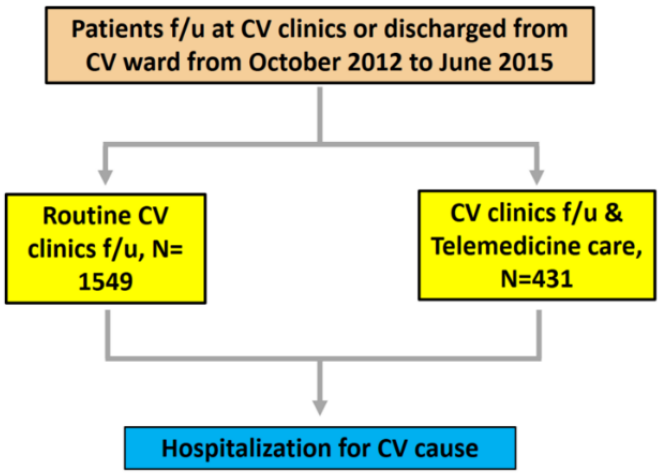


Table 1. Baseline characteristics.

Characteristics	Telehealth care group (N=431)	Control group (N=1549)	*P* value
Patient status			
Age (years), mean (SD)	70.3 (14.9)	64.2 (13.8)	<.001
Age > 80 years, n (%)	85 (19.7)	115 (7.4)	<.001
Female gender, n (%)	146 (33.9)	494 (31.9)	.45
Risk factors			
Diabetes mellitus, n (%)	128 (29.7)	369 (23.8)	.01
Hypertension, n (%)	198 (45.9)	642 (41.4)	.098
Hyperlipidemia, n (%)	143 (33.2)	484 (31.2)	.40
Atrial fibrillation, n (%)	73 (16.9)	152 (9.8)	<.001
CKD[a], n (%)	38 (8.8)	116 (7.5)	.36
CAD[b], n (%)	207 (48.0)	907 (58.6)	<.001
CHF[c], n (%)	103 (23.9)	166 (10.7)	<.001
Stroke, n (%)	64 (14.8)	128 (8.3)	<.001
PAD[d], n (%)	14 (3.2)	51 (3.3)	>.99
CHA_2DS_2-VASc[e] score, mean (SD)	2.7 (1.9)	2.0 (1.6)	<.001
Medication			
Aspirin, n (%)	184 (42.7%)	788 (50.9%)	.003
Beta-blocker, n (%)	150 (34.8%)	321 (20.7%)	<.001
ACEI/ARB[f], n (%)	170 (39.4%)	510 (32.3%)	.01
CCB[g], n (%)	44 (10.2%)	390 (25.2%)	<.001
Statin, n (%)	120 (27.8%)	440 (28.4%)	.86
OHA[h], n (%)	78 (18.1%)	229 (14.8%)	.098
Mean follow-up (days), mean (SD)	701.7 (431.4)	886.1 (530.9)	<.001

[a]CKD: chronic kidney disease.

[b]CAD: coronary artery disease.

[c]CHF: congestive heart failure.

[d]PAD: peripheral artery disease.

[e]CHA_2DS_2-VASc: congestive heart failure, hypertension, age >75 years (doubled), type 2 diabetes mellitus, previous stroke, transient ischemic attack or thromboembolism (doubled), vascular disease, age of 65-75 years, and sex.

[f]ACEI/ARB: angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker.

[g]CCB: calcium channel blocker.

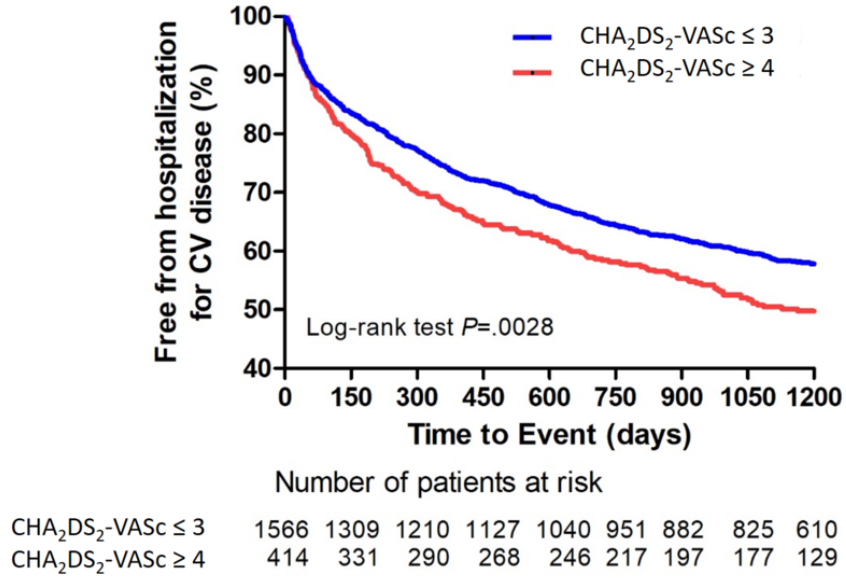
[h]OHA: oral hypoglycemic agent.

In the control group, the mean age was 64.2 (SD 13.8) years, and 68.11% were men (1055/1549). The telehealth group had more patients with congestive heart failure (103/431, 23.9% vs 166/1549, 10.7%), stroke (64/431, 14.8% vs 128/1549, 8.3%), diabetes mellitus (128/431, 29.7% vs 369/1549, 23.8%), and atrial fibrillation (73/431, 16.9% vs 152/1549, 9.8%) than the control group; all these variables were significantly different between the two groups. The CHA_2DS_2-VASc score was significantly higher in the telehealth group than in the control group (2.7 [SD 1.9] vs 2.0 [SD 1.6], *P*<.001). The mean follow-up time was 701.7 (SD 431.4) days for the telehealth group and 886.1 (SD 530.9) days for the control group (*P*<.001).

Prognosis of Patients Stratified by CHA $_2$DS$_2$-VASc Score

We stratified all patients (telehealth and control group) into higher (4-8 points) and lower (0-3 points) score groups according to their CHA_2DS_2-VASc score. A total of 414 patients were included in the higher score group and 1566 were included in the lower score group. The Kaplan-Meier curve showed significant differences in the factor remaining free of hospitalization for a cardiovascular event between the two groups. The overall estimated survival rate was 46.5% in the higher CHA_2DS_2-VASc score group and 54.8% in the lower CHA_2DS_2-VASc score group (log-rank test *P*=.003; Figure 2).

Figure 2. Kaplan-Meier curve of cardiovascular hospitalization according to CHA_2DS_2-VASc scores. The overall estimated rate of cardiovascular hospitalization was 54.8% and 46.5% in patients with CHA_2DS_2-VASc score ≤3 and ≥4, respectively (log-rank test *P*=.003). The dotted lines represented the error bars of 95% CI in both figures. CV: cardiovascular; CHA_2DS_2-VASc: congestive heart failure, hypertension, age >75 years (doubled), type 2 diabetes mellitus, previous stroke, transient ischemic attack or thromboembolism (doubled), vascular disease, age of 65-75 years, and sex.



Impact of Telehealth Program on Patients with Different CHA $_2$DS $_2$-VASc Scores

The lower CHA_2DS_2-VASc score group included 299 subjects in the telehealth group and 1267 subjects in the usual care group. The estimated survival rate of patients remaining free from cardiovascular hospitalization was 61.4% for the telehealth program and 54.3% in the control group. The Kaplan-Meier curve showed similar survival rates in both groups without significant differences (log-rank *P*=.57; Figure 3). The higher CHA_2DS_2-VASc score group included 132 subjects in the telehealth group and 282 in the usual care group. The estimated survival rate in patients remaining free from cardiovascular hospitalization was 61.5% and 41.8% in patients in the telehealth program and those in the control group, respectively. Patients accepting the telehealth program had better survival, with significant differences observed in the Kaplan-Meier curve between the two groups (log-rank *P*=.01; Figure 3).

Predictors of Cardiovascular Hospitalization for Patients With CHA_2DS_2-VASc Scores ≥4

Univariate analysis revealed a significant association between hospitalization for cardiovascular events and the following variables in patients with CHA_2DS_2-VASc scores ≥ 4 (Table 2): chronic kidney disease (hazard ratio [HR]=1.87 [CI 1.32-2.63], *P*<.001), telehealth program (HR=0.66 [CI 0.48-0.91], *P*=.01), and angiotensin receptor blocker (HR=0.70 [CI: 0.52-0.94], *P*=.006). However, only chronic kidney disease (HR=1.65 [CI 1.13-2.40], *P*=.01) and the telehealth program (HR=0.36 [CI 0.22-0.62], *P*<.001) remained significant in multivariable Cox regression. On the other hand, the variable peripheral artery disease was insignificant in univariable analysis (HR=0.97 [CI 0.74-1.29], *P*=.86) but became significant (HR=1.92 [CI 1.24-2.98], *P*=.003) after multivariable Cox regression (Table 2).

Interaction Between the Telehealth Program and the CHA_2DS_2-VASc Score

Figure 4 shows the Kaplan-Meier curve of cardiovascular hospitalization among patients with different CHA_2DS_2-VASc scores (≥4 and ≤3) who did and did not receive the fourth-generation synchronous telehealth program. In the usual care group, the overall estimated survival rate of patients who were not hospitalized for cardiovascular complications was 41.8% in group D and 54.3% in group B (Figure 4). In the telehealth group, the overall estimated survival rate of subjects remaining free from cardiovascular hospitalization was 61.5% in group C and 61.1% in group A (log-rank *P*=.0006) After a pairwise multiple comparison-adjustment procedure for Kaplan-Meier survival curve with Bonferroni correction, the estimated survival rate of group C remained significantly higher than that for group D and was similar to that of groups A and B (group A vs group C, log-rank *P*=.97; group B vs group C, log-rank *P*=.59; group C vs group D, log-rank *P*=.01; Figure 4).

Figure 3. (A) Kaplan-Meier curve of cardiovascular hospitalization and the fourth generation synchronous telehealth program in patients with CHA_2DS_2-VASc score ≤3. The overall estimated survival rate of remaining free from cardiovascular hospitalization was 61.1% in patients accepting the fourth-generation telehealth program, and 54.3% in patients not accepting the program (log-rank test *P*=.57). (B) CHA_2DS_2-VASc score ≥4. The overall estimated survival rate of remaining free from cardiovascular hospitalization was 61.5% in patients accepting the fourth-generation telehealth program and 41.8% in patients not accepting the program (log-rank test *P*=.01). CV: cardiovascular; CHA_2DS_2-VASc: congestive heart failure, hypertension, age >75 years (doubled), type 2 diabetes mellitus, previous stroke, transient ischemic attack or thromboembolism (doubled), vascular disease, age of 65-75 years, and sex.

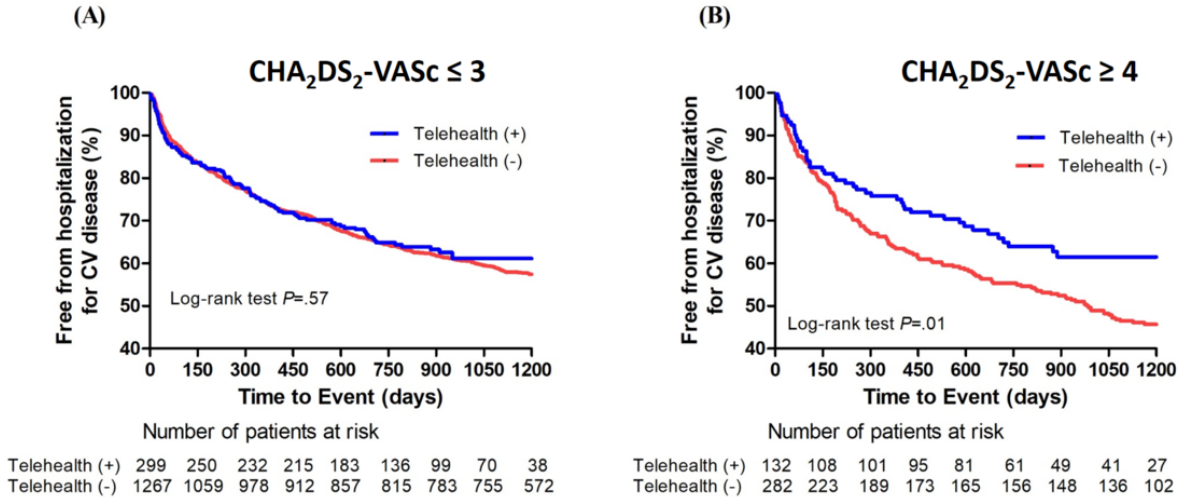
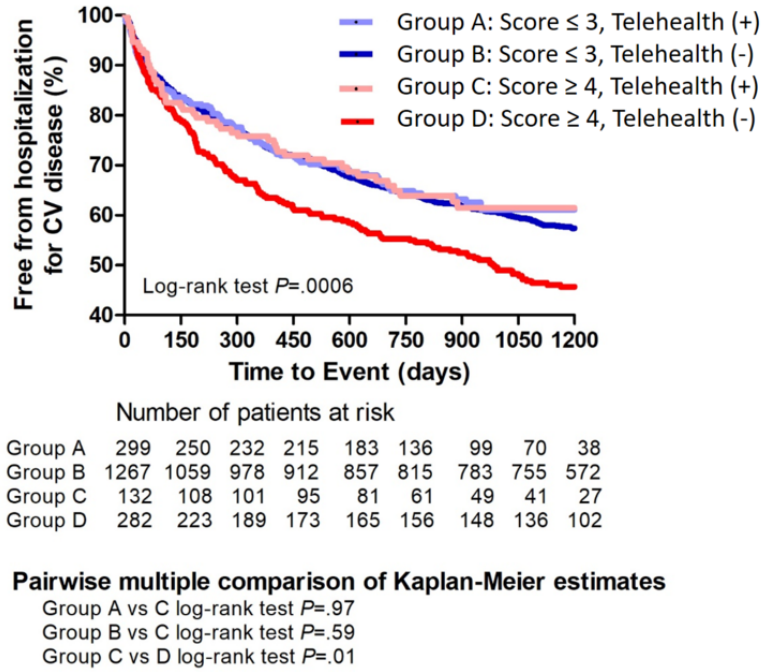


Table 2. Univariate and multivariable Cox analyses: Predictors of hospitalization for cardiovascular events in patients with CHA_2DS_2-VASc scores ≥4 (N=414).

Variables	Univariate analysis		Multivariable analysis	
	Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
Age >80 years	0.85 (0.65-1.12)	.26	0.93 (0.69-1.23)	.59
Female gender	0.86 (0.66-1.14)	.30	0.97 (0.73-1.30)	.85
Diabetes mellitus	0.99 (0.75-1.30)	.93	0.85 (0.60-1.22)	.39
Hypertension	1.05 (0.73-1.50)	.81	1.35 (0.91-2.02)	.14
Dyslipidemia	1.07 (0.82-1.41)	.62	0.94 (0.69-1.30)	.72
Coronary artery disease	1.30 (0.98-1.73)	.07	1.22 (0.88-1.68)	.23
Congestive heart failure	1.24 (0.93-1.66)	.14	1.31 (0.96-1.80)	.09
Peripheral artery disease	0.97 (0.74-1.29)	.86	1.92 (1.24-2.98)	.003
Ischemic stroke	0.76 (0.40-1.32)	.39	0.84 (0.44-1.61)	.60
Atrial fibrillation	1.03 (0.73-1.46)	.86	1.13 (0.78-1.64)	.51
Chronic kidney disease	1.87 (1.32-2.63)	<.001	1.65 (1.13-2.40)	.01
Telehealth program	0.66 (0.48-0.91)	.01	0.36 (0.22-0.62)	<.001
Antiplatelet	1.03 (0.78-1.35)	.86	1.03 (0.75-1.41)	.86
Angiotensin converting enzyme inhibitor	0.99 (0.54-1.83)	.99	0.96 (0.50-1.82)	.89
Angiotensin II receptor blocker	0.70 (0.52-0.94)	.02	0.80 (0.56-1.17)	.25
Calcium channel blocker	0.76 (0.55-1.05)	.098	0.75 (0.53-1.06)	.10
Beta-blocker	1.02 (0.74-1.40)	.92	1.06 (0.75-1.49)	.76
Statin	0.93 (0.69-1.27)	.65	1.08 (0.74-1.57)	.69
Oral hypoglycemic agent	0.75 (0.55-1.04)	.08	1.00 (0.65-1.53)	.99
Insulin	0.80 (0.44-1.47)	.47	0.97 (0.51-1.84)	.92

Figure 4. Kaplan-Meier curve of cardiovascular hospitalization in patients with different CHA_2DS_2-VASc scores (≥4 and ≤3) with/without the fourth-generation synchronous telehealth program. In the usual care group, the overall estimated survival rate of free from cardiovascular hospitalization was 41.8% in group D and 54.3% in group B. In the telehealth group, the overall estimated survival rate while remaining free of cardiovascular hospitalization was 61.5% in group C and 61.1 % in group A (log-rank test *P*=.0006). After pairwise multiple comparison adjustment procedure for Kaplan-Meier survival curve with Bonferroni correction, the estimated survival rate of group C remained significantly higher than that of group D. It was similar in groups A and B (Group A vs C: log-rank test *P*=.97; Group B vs C: log-rank test *P*=.59; Group C vs D: log-rank test *P*=.01). CV: cardiovascular; CHA_2DS_2-VASc: congestive heart failure, hypertension, age >75 years (doubled), type 2 diabetes mellitus, previous stroke, transient ischemic attack or thromboembolism (doubled), vascular disease, age of 65-75 years, and sex.



Discussion

Principal Findings

This is the first study to apply CHA_2DS_2-VASc scores to stratify patients receiving the fourth-generation synchronous telehealth program and determine who benefits most from the program. Our study showed that the fourth-generation synchronous telehealth program provided better outcomes with reduced cardiovascular hospitalization than usual care in patients with higher CHA_2DS_2-VASc scores (≥4 points).

Overview

Telehealth care has been shown to reduce hospitalizations in patients with chronic conditions such as asthma, chronic obstructive pulmonary disease, and heart failure [13-16]. We previously reported better cost effectiveness and clinical outcomes with the use of a fourth-generation synchronous telehealth program in patients with chronic cardiovascular diseases [12]. In that study, patients who received and those who did not receive telehealth programs were matched for sex, age, and Charlson score, which is a method of predicting mortality by classifying or weighting comorbid conditions [17]. However, some studies failed to show better clinical outcomes in patients receiving telehealth program [6,7]. Although the types of subjects enrolled were diverse, research on telehealth programs in chronic diseases management has shown mixed results. Development of an objective stratification system for patients and identification of which group of patients benefit from the telehealth program are needed.

The CHA_2DS_2-VASc score is a validated clinical tool predicting stroke occurrence in patients with atrial fibrillation [8,9]. Previous studies have shown that the CHA_2DS_2-VASc score enables a substantially comprehensive risk evaluation and improves physicians' ability to identify genuinely low-risk patients who have atrial fibrillation. It is not surprising that the components of these risk scores are associated with adverse outcomes because the majority of variables reflect the presence of heart disease or heart disease risk factors independent of atrial fibrillation. These grave outcomes are highlighted in studies that have examined CHA_2DS_2-VASc scores in patients without atrial fibrillation. These studies have noted a significant risk of major adverse cardiovascular events with increasing scores [18-20]. Comprehensive care for a patient with higher scores is extremely important because they have a greater risk and worse outcomes of congestive heart failure, acute coronary syndrome, and even mortality [18-20]. In this study, we did not match both groups by the Charlson score: Several components of CHA_2DS_2-VASc and Charlson scores were repetitive. Since patients with higher CHA_2DS_2-VASc scores exhibited a higher prevalence of cardiovascular diseases, the CHA_2DS_2-VASc score may be a better parameter than the Charlson score to select potential candidates for telehealth care.

In patients with nonvalvular atrial fibrillation, the $CHADS_2$ score was significantly associated with the risk of a future stroke

event. This new scoring system can identify patients with lower $CHADS_2$ scores (≤1) who remain at high risk for stroke. However, Eva et al reported that the demarcation may be minimal for patients with CHA_2DS_2-VASc scores ≤3 [21]. We apply this result to our patients in two groups, with higher (≥4) and lower (≤3) CHA_2DS_2-VASc scores. In this study, the two groups had a significant difference in survival of subjects remaining free from cardiovascular hospitalization overall. More importantly, we found that this demarcation can stratify whether patients receive benefits after accepting the telehealth program.

In the usual care group, we found that a higher CHA_2DS_2-VASc score (≥4) is significantly associated with a higher risk of cardiovascular admission compared to a lower score (≤3) (Multimedia Appendix 1). However, the cardiovascular admissions were similar for patients with CHA_2DS_2-VASc scores ≥4 and those with scores ≤3 in the telehealth group (Multimedia Appendix 1). This suggested that telehealth care could diminish the CHA_2DS_2-VASc score-associated cardiovascular admission. In the patient group with CHA_2DS_2-VASc scores ≤3, the overall estimated survival rate of subjects remaining free from cardiovascular hospitalization was similar irrespective of whether patients accepted the fourth-generation telehealth program (Figure 3). However, patients with scores ≥4 who accepted the telehealth program had significantly better outcome than those who did not accept the program (Figure 3B). After multiple comparison adjustments by Bonferroni methods, patients with scores ≥4 accepting telehealth were found to have similar clinical outcomes as patients with scores ≤3, regardless of whether they accepted the telehealth program; their outcomes were significantly better than those of patients with scores ≥4 not accepting the telehealth program (Figure 4). This implied that patients with scores ≥4 benefit most from telehealth monitoring and the need for cardiovascular hospitalization is reduced. These results showed that the CHA_2DS_2-VASc score might be a good indicator to select patients for the telehealth program.

The telehealth program has changed over time. We used the fourth-generation telehealth program, which is a synchronous and integrated remote-management program for chronic disease. This new system takes the initiative to offer an interactive environment and in-time responsiveness for patients who encountered acute illness or deterioration in condition. Compared to usual care, more accurate diagnoses and decisions can be made through quick communication after accepting the telehealth program. In our previous report, there were significantly fewer emergency department visits, hospitalizations, hospitalization days, and intensive care unit admissions per month in the telehealth group compared to the control group [12]. Thus, this new intervention program may be helpful in improving patient outcomes.

Apart from the telehealth program, we found that chronic kidney disease and peripheral artery disease were the two remaining prognostic factors for cardiovascular admission in patients with $CHAS_2DS_2$-VASc scores ≥ 4 after multivariable Cox regression. Studies on patients with chronic kidney disease accepting telehealth are rare [22,23]. We reported that among patients receiving the telehealth program, renal function status remains a predictor for first hospitalization—this is identical to our study [24]. On the other hand, patients with peripheral artery disease accepting the telehealth program are rarer despite their higher mortality rate [25]. According to previous studies, the mortality of symptomatic and asymptomatic patients was 19% and 24%, respectively, at 5 years [26]. Furthermore, patients with peripheral artery disease share similar risk factors as patients with coronary artery disease, carotid artery stenosis, and congestive heart failure as compared to the general population [27-29]. Early identification of peripheral artery disease with optimized and comprehensive treatment is mandatory to improve clinical outcomes; the telehealth program may have a role in this process. There have been studies addressing the issue of telehealth applications in patients with peripheral artery disease with wounds or gangrene, but these studies focused on wound infection control rather than general patient care [30,31]. Larger clinical trials applying telehealth care to patients with peripheral artery disease may be needed to improve clinical outcomes.

Limitations

This study has several limitations. First, the study was not randomized, which resulted in heterogeneity of the patient population, disease severity, and patient selection. Patients with the capacity to self-monitor and be enrolled in a program for home remote monitoring are potentially more likely to receive medical assistance when experiencing clinical changes or a clinical decline. The patient-selection process should be recognized as a limitation. Second, the presence of numerous confounding factors in our cohort might have influenced the result, including the missing events. We tried to perform multivariable Cox regression analysis to minimize the possible confounding effect of other clinical factors. Third, there might be some statistical limitations and considerations of statistical testing/modeling. For example, peripheral artery disease was not significantly associated with hospitalization for cardiovascular events in the univariate Cox regression but became significant in the multivariable Cox model in patients with CHA_2DS_2-VASc scores ≥4 (Table 2). This might be due to some reverse confounding or overfitting. On the other hand, the survival analysis stratified by CHA_2DS_2-VASc scores (≤3 vs >4) may induce type I error inflation (Figure 3). However, the result would be significant even if Bonferroni correction was done in patients with CHA_2DS_2-VASc scores ≥4. This indicates that type I error inflation may not be a serious problem in this study. Fourth, the clinical outcomes were derived from the electronic billing and medical records of our hospital, and the patients who received care outside our hospital were not recorded. Resources that were used but not billed may have also been overlooked when extracting data from our billing system.

Conclusions

Patients with higher CHA_2DS_2-VASc scores had higher risks of cardiovascular admissions, but the fourth-generation telehealth program could diminish the outcome difference associated with scores. Patients with CHA_2DS_2-VASc scores ≥4 benefited the most from the fourth-generation telehealth program and remained free of cardiovascular hospitalization.

Acknowledgments

We thank Kuo-Liong Chien for his consultation in the statistical analysis during data analysis. This research was supported by the Joint Research Center for AI Technology and All Vista Healthcare under Ministry of Science and Technology of Taiwan (MOST 107-2634-F-002-018) and Center for Artificial Intelligence & Advanced Robotics, and National Taiwan University Hospital (106-A136, 107-EDN11, 108-N4406).

Conflicts of Interest

None declared.

Multimedia Appendix 1

(A) Kaplan-Mayer curve of cardiovascular hospitalization with CHA_2DS_2-VASc score in patients who did not accept the fourth-generation synchronous telehealth program. The overall estimated survival rate of cardiovascular hospitalization was 41.8% in patients with CHA_2DS_2-VASc score ≥4 and 54.3% in patients with a score ≤3 (log-rank test *P*=.0001). (B) Patients using the fourth-generation synchronous telehealth program. The overall estimated survival rate of cardiovascular hospitalization was 61.5% in patients with CHA_2DS_2-VASc score ≥4 and 61.1% in patients with the score ≤3 (log-rank test *P*=.97). CV: cardiovascular; CHA_2DS_2-VASc: congestive heart failure, hypertension, age > 75 years (doubled), type 2 diabetes mellitus, previous stroke, transient ischemic attack or thromboembolism (doubled), vascular disease, age of 65-75 years, and sex.

[PNG File, 311KB - jmir_v21i1e12790_app1.png]

References

1. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. J Am Coll Cardiol 2017 Jul 04;70(1):1-25 [FREE Full text] [doi: 10.1016/j.jacc.2017.04.052] [Medline: 28527533]
2. Vernooij J, Kaasjager H, van der Graaf Y, Wierdsma J, Grandjean H, Hovens M, SMARTStudy Group. Internet based vascular risk factor management for patients with clinically manifest vascular disease: randomised controlled trial. BMJ 2012 Jun 12;344:e3750 [FREE Full text] [doi: 10.1136/bmj.e3750] [Medline: 22692651]
3. Kitsiou S, Paré G, Jaana M. Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015 Mar 12;17(3):e63 [FREE Full text] [doi: 10.2196/jmir.4174] [Medline: 25768664]
4. Inglis S, Clark R, McAlister F, Stewart S, Cleland J. Which components of heart failure programmes are effective? A systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: Abridged Cochrane Review. Eur J Heart Fail 2011 Sep;13(9):1028-1040 [FREE Full text] [doi: 10.1093/eurjhf/hfr039] [Medline: 21733889]
5. Hung C, Yu J, Lin Y, Chen Y, Huang C, Lee J, et al. Mortality Benefit of a Fourth-Generation Synchronous Telehealth Program for the Management of Chronic Cardiovascular Disease: A Longitudinal Study. J Med Internet Res 2016 May 13;18(5):e102 [FREE Full text] [doi: 10.2196/jmir.5718] [Medline: 27177497]
6. Takahashi PY, Pecina JL, Upatising B, Chaudhry R, Shah ND, Van Houten H, et al. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. Arch Intern Med 2012 May 28;172(10):773-779 [FREE Full text] [doi: 10.1001/archinternmed.2012.256] [Medline: 22507696]
7. Wootton R. Twenty years of telemedicine in chronic disease management--an evidence synthesis. J Telemed Telecare 2012 Jun;18(4):211-220 [FREE Full text] [doi: 10.1258/jtt.2012.120219] [Medline: 22674020]
8. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, ESC Scientific Document Group. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. Eur Heart J 2016 Dec 07;37(38):2893-2962. [doi: 10.1093/eurheartj/ehw210] [Medline: 27567408]
9. January CT, Wann LS, Alpert JS, Calkins H, Cigarroa JE, Cleveland JC, ACC/AHA Task Force Members. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the Heart Rhythm Society. Circulation 2014 Dec 02;130(23):2071-2104. [doi: 10.1161/CIR.0000000000000040] [Medline: 24682348]
10. Mitchell LB, Southern DA, Galbraith D, Ghali WA, Knudtson M, Wilton SB, APPROACH investigators. Prediction of stroke or TIA in patients without atrial fibrillation using CHADS2 and CHA2DS2-VASc scores. Heart 2014 Oct;100(19):1524-1530. [doi: 10.1136/heartjnl-2013-305303] [Medline: 24860007]
11. Mazzone C, Cioffi G, Carriere C, Barbati G, Faganello G, Russo G, et al. Predictive role of CHADS-VASc score for cardiovascular events and death in patients with arterial hypertension and stable sinus rhythm. Eur J Prev Cardiol 2017 Dec;24(15):1584-1593. [doi: 10.1177/2047487317726068] [Medline: 28812917]

12. Ho Y, Yu J, Lin Y, Chen Y, Huang C, Hsu T, et al. Assessment of the cost-effectiveness and clinical outcomes of a fourth-generation synchronous telehealth program for the management of chronic cardiovascular disease. J Med Internet Res 2014 Jun 10;16(6):e145 [FREE Full text] [doi: 10.2196/jmir.3346] [Medline: 24915187]
13. McLean S, Chandler D, Nurmatov U, Liu J, Pagliari C, Car J, et al. Telehealthcare for asthma. Cochrane Database Syst Rev 2010 Oct 06(10):CD007717. [doi: 10.1002/14651858.CD007717.pub2] [Medline: 20927763]
14. McLean S, Nurmatov U, Liu J, Pagliari C, Car J, Sheikh A. Telehealthcare for chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2011 Jul 06(7):CD007718. [doi: 10.1002/14651858.CD007718.pub2] [Medline: 21735417]
15. Clark R, Inglis S, McAlister F, Cleland J, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. BMJ 2007 May 05;334(7600):942 [FREE Full text] [doi: 10.1136/bmj.39156.536968.55] [Medline: 17426062]
16. Clarke M, Shah A, Sharma U. Systematic review of studies on telemonitoring of patients with congestive heart failure: a meta-analysis. J Telemed Telecare 2011;17(1):7-14. [doi: 10.1258/jtt.2010.100113] [Medline: 21097564]
17. Charlson M, Pompei P, Ales K, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373-383. [Medline: 3558716]
18. Crandall M, Horne B, Day J, Anderson J, Muhlestein J, Crandall B, et al. Atrial fibrillation significantly increases total mortality and stroke risk beyond that conveyed by the CHADS2 risk factors. Pacing Clin Electrophysiol 2009 Aug;32(8):981-986. [doi: 10.1111/j.1540-8159.2009.02427.x] [Medline: 19659615]
19. Svendsen J, Nielsen J, Darkner S, Jensen G, Mortensen L, Andersen HR, DANPACE Investigators. CHADS2 and CHA2DS2-VASc score to assess risk of stroke and death in patients paced for sick sinus syndrome. Heart 2013 Jun;99(12):843-848 [FREE Full text] [doi: 10.1136/heartjnl-2013-303695] [Medline: 23539553]
20. Ntaios G, Lip G, Makaritsis K, Papavasileiou V, Vemmou A, Koroboki E, et al. $CHADS_2$, CHA_2 S_2 DS_2 -VASc, and long-term stroke outcome in patients without atrial fibrillation. Neurology 2013 Mar 12;80(11):1009-1017. [doi: 10.1212/WNL.0b013e318287281b] [Medline: 23408865]
21. Pueo E, Campos B, Anguita M, Worner F. Does CHA_2 DS_2 -VASc score select patients who will benefit most from anticoagulation? Rev Esp Cardiol (Engl Ed) 2014 May;67(5):417-418. [doi: 10.1016/j.rec.2014.01.005] [Medline: 24774741]
22. Ishani A, Christopher J, Palmer D, Otterness S, Clothier B, Nugent S, Center for Innovative Kidney Care. Telehealth by an Interprofessional Team in Patients With CKD: A Randomized Controlled Trial. Am J Kidney Dis 2016 Jul;68(1):41-49. [doi: 10.1053/j.ajkd.2016.01.018] [Medline: 26947216]
23. Ong S, Jassal S, Miller J, Porter E, Cafazzo J, Seto E, et al. Integrating a Smartphone-Based Self-Management System into Usual Care of Advanced CKD. Clin J Am Soc Nephrol 2016 Dec 06;11(6):1054-1062 [FREE Full text] [doi: 10.2215/CJN.10681015] [Medline: 27173169]
24. Hung C, Lee J, Chen Y, Huang C, Wu V, Wu H, et al. Effect of Contract Compliance Rate to a Fourth-Generation Telehealth Program on the Risk of Hospitalization in Patients With Chronic Kidney Disease: Retrospective Cohort Study. J Med Internet Res 2018 Jan 24;20(1):e23 [FREE Full text] [doi: 10.2196/jmir.8914] [Medline: 29367185]
25. Davins Riu M, Borràs Pérez X, Artigas Raventós V, Palomera Fanegas E, Serra Prat M, Alós Villacrosa J. Use of Telehealth as a New Model for Following Intermittent Claudication and Promoting Patient Expertise. Telemed J E Health 2018 Oct;24(10):773-781. [doi: 10.1089/tmj.2017.0226] [Medline: 29323628]
26. Diehm C, Allenberg J, Pittrow D, Mahn M, Tepohl G, Haberl R, German Epidemiological Trial on Ankle Brachial Index Study Group. Mortality and vascular morbidity in older adults with asymptomatic versus symptomatic peripheral artery disease. Circulation 2009 Nov 24;120(21):2053-2061. [doi: 10.1161/CIRCULATIONAHA.109.865600] [Medline: 19901192]
27. Sukhija R, Yalamanchili K, Aronow W. Prevalence of left main coronary artery disease, of three- or four-vessel coronary artery disease, and of obstructive coronary artery disease in patients with and without peripheral arterial disease undergoing coronary angiography for suspected coronary artery disease. Am J Cardiol 2003 Aug 01;92(3):304-305. [Medline: 12888139]
28. Ness J, Aronow W, Newkirk E, McDanel D. Prevalence of symptomatic peripheral arterial disease, modifiable risk factors, and appropriate use of drugs in the treatment of peripheral arterial disease in older persons seen in a university general medicine clinic. J Gerontol A Biol Sci Med Sci 2005 Feb;60(2):255-257. [Medline: 15814871]
29. Wu C, Lee J, Chiang F, Yang C, Huang S, Hwang J, et al. Plasma levels of tumor necrosis factor-α and interleukin-6 are associated with diastolic heart failure through downregulation of sarcoplasmic reticulum Ca2+ ATPase. Crit Care Med 2011 May;39(5):984-992. [doi: 10.1097/CCM.0b013e31820a91b9] [Medline: 21263314]
30. van Netten JJ, Clark D, Lazzarini P, Janda M, Reed L. The validity and reliability of remote diabetic foot ulcer assessment using mobile phone images. Sci Rep 2017 Aug 25;7(1):9480 [FREE Full text] [doi: 10.1038/s41598-017-09828-4] [Medline: 28842686]
31. Smith-Strøm H, Igland J, Østbye T, Tell G, Hausken M, Graue M, et al. The Effect of Telemedicine Follow-up Care on Diabetes-Related Foot Ulcers: A Cluster-Randomized Controlled Noninferiority Trial. Diabetes Care 2018 Dec;41(1):96-103. [doi: 10.2337/dc17-1025] [Medline: 29187423]

Abbreviations

ACEI: angiotensin converting enzyme inhibitor

ARB: angiotensin II receptor blocker
CAD: coronary artery disease
CCB: calcium channel blocker
CHA_2DS_2-VASc: congestive heart failure, hypertension, age >75 years (doubled), type 2 diabetes mellitus, previous stroke, transient ischemic attack or thromboembolism (doubled), vascular disease, age of 65-75 years, and sex
CHF: congestive heart failure
CKD: chronic kidney disease
HR: hazard ratio
OHA: oral hypoglycemic agent
PAD: peripheral artery disease

Edited by G Eysenbach; submitted 11.11.18; peer-reviewed by M Singh, I Kedan, X Shen, K Mann; comments to author 03.01.19; revised version received 08.01.19; accepted 13.01.19; published 31.01.19.

Please cite as:
Lee JK, Hung CS, Huang CC, Chen YH, Chuang PY, Yu JY, Ho YL
Use of the CHA2DS2-VASc Score for Risk Stratification of Hospital Admissions Among Patients With Cardiovascular Diseases Receiving a Fourth-Generation Synchronous Telehealth Program: Retrospective Cohort Study
J Med Internet Res 2019;21(1):e12790
URL: https://www.jmir.org/2019/1/e12790/
doi:10.2196/12790
PMID:30702437

Original Paper

A Web-Based Application to Improve Data Collection in an Interventional Study Targeting Childhood Obesity: Pre-Post Analysis

Meagan M Hanbury[1], PhD; Banafsheh Sadeghi[2], MD, PhD; Iraklis Erik Tseregounis[3], PhD; Rosa Gomez-Camacho[4], PhD; Rosa D Manzo[5], PhD; Maria Isabel Rangel[6], MPH; Bogdan Alexandrescu[1], PhD; Adela de la Torre[7], PhD

[1]Center for Transnational Health, University of California, Davis, Davis, CA, United States
[2]Department of Internal Medicine, School of Medicine, University of California, Davis, Davis, CA, United States
[3]Center for Healthcare Policy and Research, University of California, Davis, Davis, CA, United States
[4]Office of Planning & Institutional Performance, Florida Gulf Coast University, Fort Myers, FL, United States
[5]Health Science Research Institute, University of California, Merced, Merced, CA, United States
[6]School of Public Health, University of California, Berkeley, Berkeley, CA, United States
[7]San Diego State University, San Diego, CA, United States

Corresponding Author:
Meagan M Hanbury, PhD
Center for Transnational Health
University of California, Davis
328 D Street
Davis, CA, 95616
United States
Phone: 1 617 877 6795
Email: mhanbury@ucdavis.edu

Abstract

Background: Although participatory action research (PAR) studies have proliferated in recent years, the development of technological resources to manage these types of projects has not kept pace. Few studies show how Web-based applications can be used to efficiently manage the data collection process.

Objective: This study described the development, use, and impact of a Web-based application to facilitate data management in Niños Sanos, Familia Sana (Healthy Children, Healthy Family), an interventional multifaceted PAR field study.

Methods: We described the transformation of the data management process and evaluated the impact of the application in terms of time efficiency of data collection and engagement of community-based data collectors. We defined time efficiency as the total number of days it took to collect 3 main surveys, per year of data collection. The engagement of data collectors was assessed based on qualitative reports.

Results: The amount of time it took to perform a round of data collection was reduced after implementation of the field team application (between 382 and 383 days and 198 and 233 days). Secondary data were also collected in a tighter time frame around collection of the primary outcome, and communication among data collectors, the field staff, and the research team was streamlined. In focus groups, community-based data collectors reported feeling more empowered and engaged in the data collection process after implementation of the application.

Conclusions: A Web-based management application was successful in improving data collection time efficiency and engagement among data collectors.

(J Med Internet Res 2019;21(1):e10861) doi:10.2196/10861

KEYWORDS

data collection; internet; rural population; efficiency

Introduction

Participatory action research (PAR) methodologies have been employed in the public health field as a best practice to develop interventions that aim to address ethnic and socioeconomic disparities in health services [1,2]. PAR studies are often conducted in populations where participants are hard to reach and retain, requiring an iterative process of data collection methods and a high degree of coordination among researchers and community staff [3]. Experts have suggested that establishing a field office and hiring local community members as staff are key components to a successful community-university partnership [4]. Nonetheless, the geographical distance and skill differences between researchers and community staff can create challenges such as the following: communication lags between the research team and field staff, limited exposure of field staff to university-based applied research methods, and undervaluing community members' insights about translating research design elements of the project into appropriate community-based methods and strategies.

Although PAR studies have proliferated in recent years, the development of technological resources to manage these types of projects has not kept pace. Computers, smartphones, and other mobile computing devices have increasingly been used to collect data [5]. However, few studies show how mobile and Web-based technologies can be used to efficiently manage the data collection process. Oftentimes, data collection and participant information are managed in an offline database that does not allow for real-time updates. Studies that have employed mobile and Web-based technology as a management tool report mixed results. Zhang et al found that using mobile device–based technology improved the validity and reliability of data, and the availability of real-time communication was beneficial for project directors and data collectors [6]. However, Vallieres et al found that there was no change in perceived supervision, engagement, or motivation among community health workers after the implementation of a mobile health app as a human resource tool [7]. Vallieres et al suggested that differences in outcomes may hinge on the engagement and motivation of the people using the technology.

This study described the development, use, and impact of a Web-based application to facilitate data management in *Niños Sanos, Familia Sana* (NSFS; translated as Healthy Children, Healthy Family), an interventional multifaceted PAR field study. NSFS evaluated the impact of an intervention package on the rate of growth of body mass index (BMI) among children of Mexican heritage in California's Central Valley. A number of logistical challenges emerged through the initial data collection phase—largely due to the geographical remoteness of the field site and hard-to-reach nature of the participant communities. In response to these challenges, researchers and field staff worked with a computer programmer to create a Web-based data management tool. The Web-based data management tool was not included in the original study protocol and was developed, through a learning process, as a solution to challenges associated with the initial data collection methods. Our goal was to describe the transformation of the data management process and evaluate the impact of the application in terms of time efficiency of data collection and engagement of community-based data collectors. We defined time efficiency as the total number of days it took to collect 3 main surveys, per year of data collection. The engagement of data collectors was assessed based on qualitative reports.

Methods

Overview

NSFS was a multifaceted, public health intervention study designed to slow the rate of growth of BMI among children of Mexican heritage aged between 3 and 8 years. Recruitment and enrollment began in August 2011. The intervention phase occurred from September 2012 through August 2015 and follow-up concluded in April 2016. Throughout the study, 782 children were recruited and enrolled. The methodology and study protocol have been described elsewhere [8]. NSFS was approved by the University of California (UC) Davis Institutional Review Board, and the legal guardians of eligible children provided written informed consent.

NSFS was located within 5 rural communities in California's Central Valley. These communities are 160 to 190 miles from UC Davis, the institution that hosted the research project. The geographic distribution of these communities is dispersed, and the distance between the communities ranges from 5 to 30 miles. The location and geographical dispersion of the field sites created challenges to data collection, including long travel times and difficulty for the university researchers to be physically engaged in all aspects of data collection. Furthermore, the demographic and economic makeup of the study population contributed additional logistical challenges. Many of the parents and guardians are employed in agricultural professions and work long, variable hours making scheduling of data collection difficult. Households also return to Mexico and spend extended periods of time out of the country. Data were collected via verbal interviews, as opposed to Web-based surveys, as many households have limited access and exposure to technology as well as a low educational level. A number of parents and guardians speak only Spanish, requiring a culturally nuanced data collection team.

The implementation and evaluation of NSFS was a coordinated effort among researchers at UC Davis, public health professionals from UC Cooperative Extension (UCCE), a field team of local staff, community health workers (promotoras), and local students from West Hills Community College (WHCC). The UCCE team members were based in Fresno, a city 30 to 40 miles from the study communities. Given the scope of the study and the distance of the field site to UC Davis, a field office was established in the intervention community. The field office was staffed by a local study coordinator, local program representative, and a local nutrition educator. The study coordinator supervised a team of data collectors that included promotoras and WHCC students.

Data Collection

Throughout the NSFS study, numerous types of data were collected from eligible children and their households. The primary outcome of the study was the rate of growth in child

BMI. Anthropometric measurements were collected from eligible children at baseline and every 6 months thereafter. Additional data were collected throughout the study to both assess secondary outcomes and understand factors that influence behavioral choices. Data sources included biannual anthropometric measurements, annual surveys, monthly nutrition class attendance records, biannual accelerometers, ongoing loyalty card data from grocery store scanners, annual food shopping receipts, and biannual carotenoid measurements.

Data collection occurred at the field office, community events, school sites, and participants' homes. In the first 2 years of data collection, the study coordinator would contact parents and guardians and schedule both the time and location of a data collection session. The study coordinator would also assign a data collector to the scheduled session. Surveys were collected using LimeSurvey, an open source Web application (Version 1.91), via remote Netbooks (Gateway LT2000). Data were uploaded from the Netbooks through an internet connection to a secure MySQL (Oracle) database, cloud-hosted on a private server.

Initial Tracking System

During the recruitment and enrollment period, a master list was developed to track basic household information, eligibility status, and willingness to participate for all recruited households. Upon enrollment, households were assigned a unique household identification number for tracking and confidentiality purposes. Individuals within the household were also assigned an individual identification number. All tracking numbers were included in the master list. The master list was managed in a Microsoft Excel spreadsheet by the field staff.

An activity tracking log was developed in concert with the master list to monitor intervention and data collection activities among households. The activity tracking log was also managed in Microsoft Excel by the field staff. Before a scheduled data collection session, the field staff would use the activity tracking log to prepare a report that included basic household information and a list of data elements that needed to be collected. Reports would be printed and given to the data collector assigned to the session. Data collectors would update the report with any new household information and a list of surveys that had been collected at the session. They would then return the report to the field staff for entry into the activity log.

Challenges

The initial system of data management had a number of drawbacks. Updating the activity log and creating reports was an inefficient process. It required a great deal of coordination and communication among the field staff, data collectors, and the research team. Survey progress and household information had to be entered twice—once on paper and again in the electronic log. Data collectors occasionally lost or did not return paper reports in a timely manner. Other times data collection activities were not recorded in the electronic log by the field staff. As a result, households were sometimes asked to complete surveys more than once. To ensure data collection was on track, researchers made weekly trips to the field site to reconcile the activity log with data that had been uploaded to the database.

Scheduling of data collection activities was also inefficient. The field staff needed to coordinate with both households and data collectors to find a mutually agreeable time and location for data collection activities. In some cases, households did not have time to complete surveys in one sitting or had a schedule change and were not available at the time of their appointment. The field staff would have to reschedule both the data collector and household, which further delayed data collection and increased travel time and cost.

Furthermore, the initial system of data management yielded a disconnect among the data collectors, the field staff, and the research team. Data collectors would call the field office to report issues with surveys or specific households, and the field staff would contact the research team for guidance. This was a time-consuming way to troubleshoot problems that usually required quick implementation of solutions. Researchers also lacked the ability to monitor data collection activities in real time and offer appropriate feedback. For example, the Food Frequency Questionnaire was only required from 1 predefined child per family and in some cases the wrong child was interviewed. Conversely, data collectors had a wealth of knowledge about the community and household-specific constraints (such as work schedules) that was not easily communicated back to the research team and field staff nor was it considered in scheduling decisions.

Development of Field Team Application

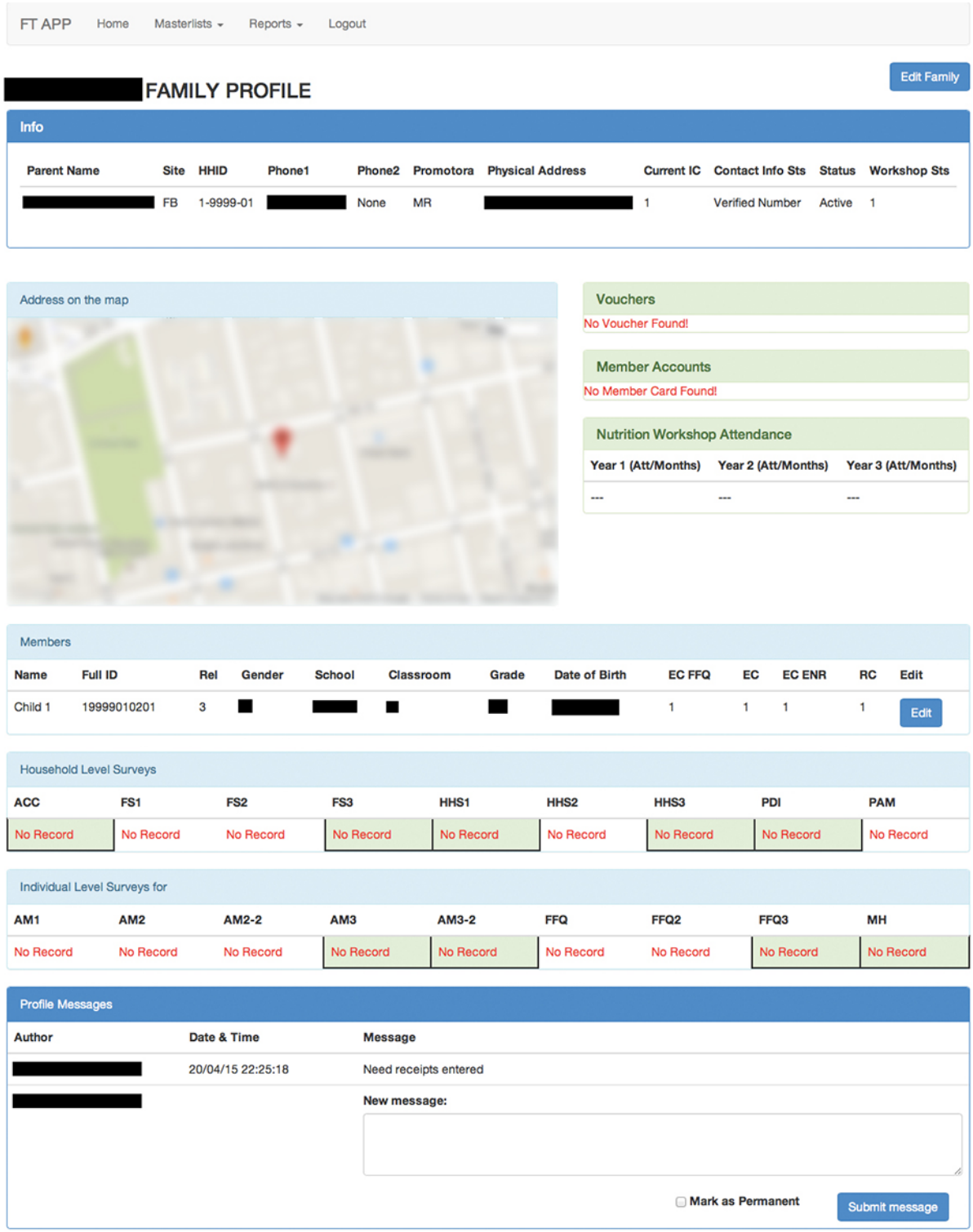
To simplify oversight of data collection, the researchers, field staff, and data collectors worked with a computer programmer to develop a field team application. The field team application is a secure, password-protected website application that tracks household and individual information as well as data collection progress. The application contains searchable master lists at both the household and individual level with high-level information including name, identification number, contact information, and eligibility and participation status. Each household and individual entry also links to a full household profile page. The household profile contains information from the master list, as well as detailed information about each member of the household, status of household and individual-level surveys and instruments, a map indicating the household's location, participation records for intervention components, and a messaging portal (see Figure 1 for an example of the profile page). Information can be filtered by a number of variables, such as status of a particular survey, to provide summary reports.

The process of building the application took several months of interactive meetings among stakeholders to understand the data collection process and translate management to a Web-based portal. As a result, the application is tailored to the specifics of the NSFS study. The computer programmer also worked with the field staff and data collectors to ensure the application was user friendly to an audience with varied technological literacy. An initial version of the application was tested before rollout. The application was officially launched on September 1, 2014, at the start of the third year of data collection.

Following implementation of the field team application, all data collection activities were tracked directly within the application.

The application was linked to the MySQL database. Data collection progress was updated within the application in real time as data elements were uploaded from Netbooks to the database. Scheduling procedures also shifted after implementation. The study coordinator assigned data collectors to specific households based on previous working relationships. Data collectors took ownership of these households and scheduled collection activities directly with participants. The researchers and field staff were able to take a more hands-off approach to management and monitored data collection remotely. The messaging section allowed data collectors to communicate efficiently with the field staff and research team. Data collectors were also able to take notes within the application and include supplementary information about households in the profile page. Overall, there was greater transparency of the data collection process among all stakeholders after implementation of the field team application.

Figure 1. Example profile page.



Field Team Application Evaluation Methodology

Quantitative: We assessed time efficiency of data collection by comparing the total number of days it took to collect 3 main surveys in years 2, 3, and 4. Apart from the implementation of the field team application, which occurred after year 2 collection and before year 3 collection, the survey instruments and procedures remained largely the same throughout.

Qualitative: As part of evaluating the community-based aspects of the NSFS study, we conducted focus groups with data collectors in one of the study sites. Focus group methodology has been described as a useful methodology to allow participants to engage in a free-flowing discussion and to express their attitudes and opinion about a specific focus [9,10]. The focus groups were facilitated in Spanish by 2 bilingual, bicultural researchers and a project coordinator. The focus groups lasted between 1.5 and 2 hours and were recorded using a digital voice recorder. A bilingual researcher transcribed the focus group discussions, and a second researcher reviewed the transcription. The coding of the data followed a deductive approach based on Strauss's (1987) methodology [11,12].

Results

This study assessed changes associated with the implementation of field team application in 2 main areas: time efficiency of data collection and engagement of the data collectors. Figure 2 shows the probability density of data collection over time for 3 main surveys and reflects the change in time efficiency before and after the implementation of the application. In year 2, surveys took between 382 and 383 days to collect for the whole sample. In years 3 and 4, following application implementation, collection took between 198 and 233 days. Half of the sample was surveyed in under 85 days for the years following application implementation, whereas collection for half of the sample in the year before ranged between 111 and 113 days.

The main outcome measure of the study is the change in BMI in an eligible child over time. The supporting surveys were conducted to provide an analytical context for modeling the main outcome. To maximize their explanatory value, a tight time frame of survey collection around anthropometric measures was crucial. Figure 3 shows the distribution of time elapsed between collection of anthropometric measurements and supporting surveys. Half of the surveys were collected at the 190-day mark following anthropometric measurements in year 2. In year 3, this had improved to 70 days (a 63% decrease) and by year 4 to 55 days (a 71% decrease).

Figure 2. Progression of survey collection.

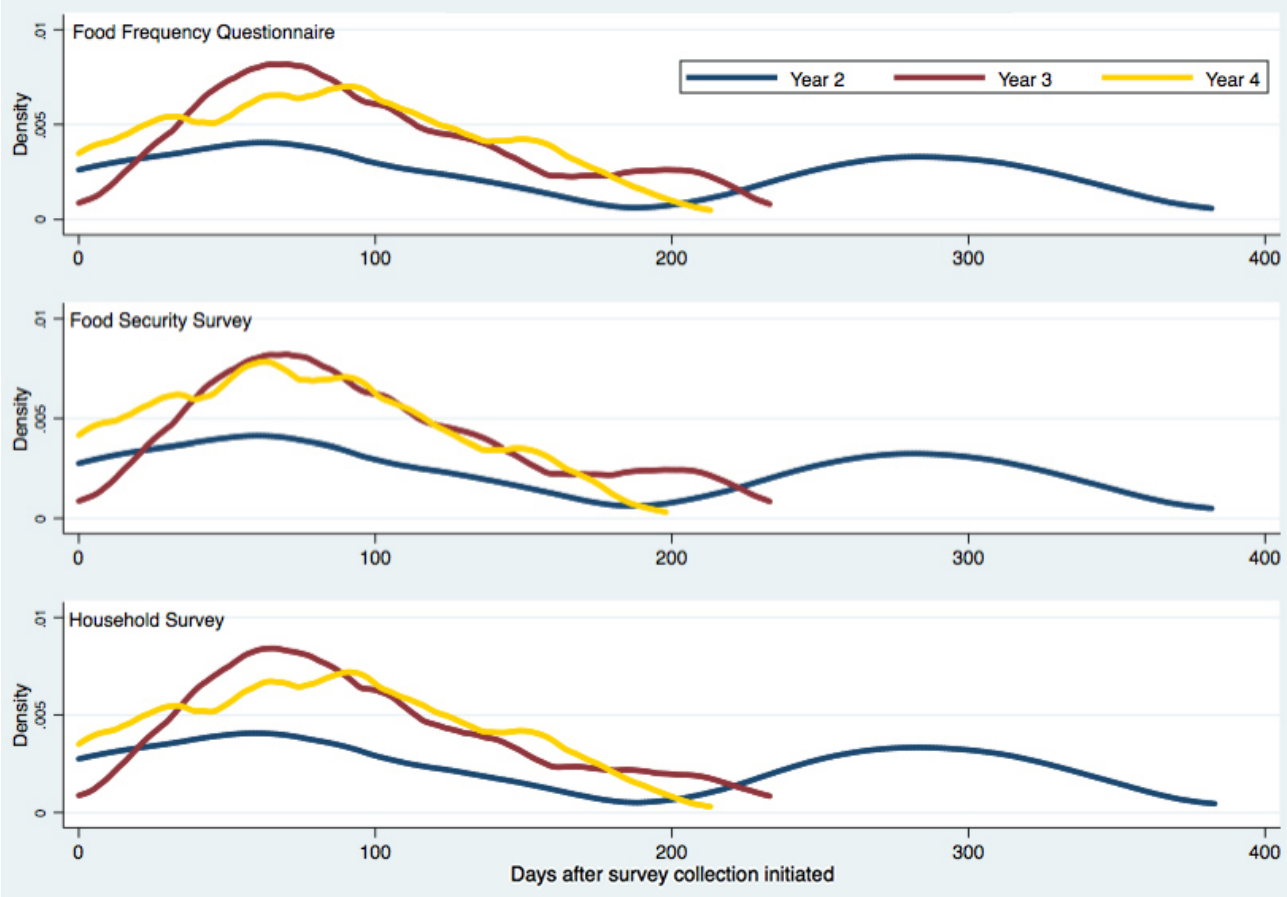
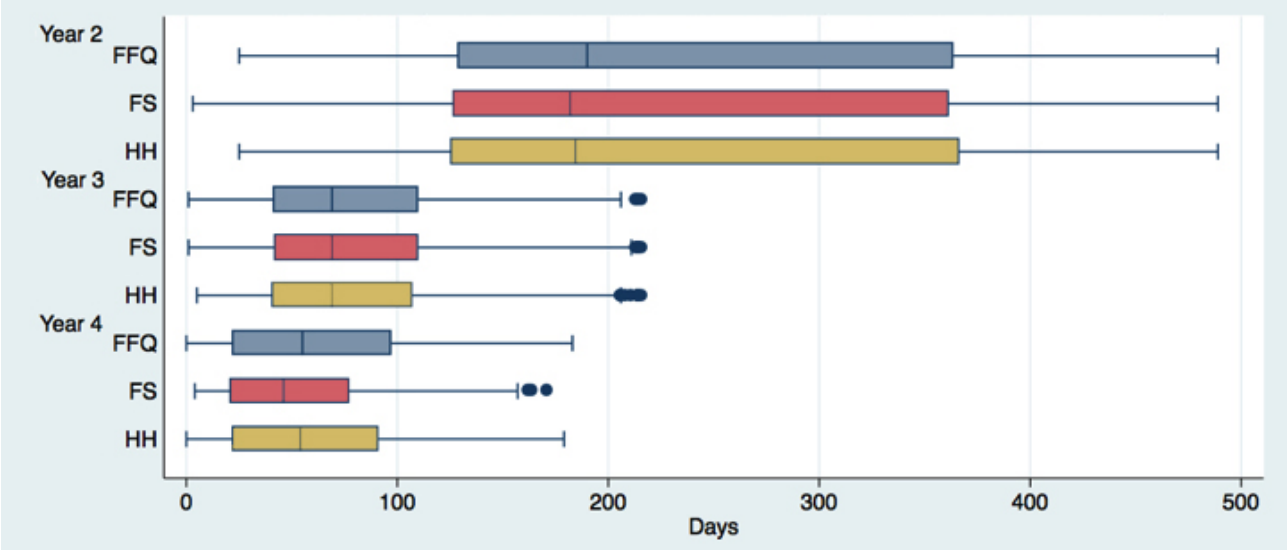


Figure 3. Time between medical measurements and surveys. FFQ: Food Frequency Questionnaire; FS: Food Security Survey; HH: Household Survey.



In addition to improvements in time efficiency, data collectors became more engaged in the collection process after implementation of the field team application. Overall, 9 data collectors participated in 2 focus groups. All of the data collectors were women of Mexican origin. The data collectors' age ranged between 19 and 52 years. During focus group discussions [9], data collectors shared their experience with the field team application and reported being able to communicate in an efficient and direct manner with the research team. For example, 1 data collector mentioned:

> *When we had the application, we were able to input comments directly into the comment box, which allowed the university team [researchers] to see the comment. That saved us a lot of time since we did not have to communicate with [the study coordinator], and then wait for her to deliver the message to someone else.*

This example describes the data collectors' experiences of having the ability to communicate directly with the research team, which created a sense of belongingness to the research team and empowerment. More specifically, the data collectors mentioned:

> *Having the ability to collect data and communicate with the research team made us feel as part of the team.*

Additionally, the comment feature of the application allowed the data collectors to update the research team on family-specific problems or challenges, which in turn facilitated the overall tracking of participants. For example, the data collectors indicated that if a family had communicated with them that they would be out of town or working seasonal jobs outside the community, then the data collectors could input that information in the comment box. Having that information readily available in the application allowed them to plan their strategy for contacting families.

Finally, the data collectors expressed that the application allowed them to be more effective in working as a team because it allowed them to allocate responsibilities among themselves and work collaboratively. For example, 1 data collector described:

> *The application allowed me to see how many families [other data collectors] had completed or if there was a specific issue in scheduling one family and I knew that family personally or was going to be around the area where they live, then I would follow up or coordinate with one of the other data collectors to find the best way to collect the data.*

Discussion

Principal Findings

A centralized, Web-based information portal improved data collection and facilitated communication in a complex public health field study. Results show that the amount of time it took to perform a round of data collection was reduced after implementation of the field team application. Secondary data were also collected in a tighter time frame around collection of the primary outcome, increasing the explanatory value of these complementary surveys. Additionally, communication among data collectors, the field staff, and the research team was streamlined. Community-based data collectors were also provided greater technological skills that further empowered their input in the data management process.

The NSFS study experienced data management challenges commonly faced by PAR projects. We had numerous data elements and a hard-to-reach population. Many of the parents involved in this project were employed in the agricultural sector and worked long, variable hours. Furthermore, some families spent extended amounts of time out of the country, and others lacked transportation to attend community-wide data collection events or office visits. The rural nature of the community meant data collectors spent considerable time on transportation, particularly if they had to make repeated visits to a participant's home. Our data collection methods were revised and updated over time to respond to these challenges; however, the geographic distance between the field site and university made

real-time communication among researchers, field staff, and data collectors difficult.

The field team application and resulting improvements in data management were critical to the success of this project. Our research protocol required yearly data collection to assess changes in child BMI over time. Implementation of the field team application enabled us to meet this data collection goal. Improvements in time efficiency largely occurred because fewer transactions among researchers, the field staff, data collectors, and participants were needed to complete data collection. We were able to eliminate the time lag created by paper reports and an offline management system. Data collectors were able to schedule and perform data collection directly with remote oversight from the field staff and researchers. Data collectors had intimate knowledge of the community that allowed them to target families at convenient times and locations, as well as to work as a team to reduce transportation time and repeated visits. With greater technological efficiency, the research team was able to fully benefit from the knowledge of the community-based data collectors. Remote, real-time oversight also allowed the research team to course-correct quickly when needed.

The field team application contributed to the success of this study in other, less measurable ways as well. The application reduced the management burden associated with data collection and allowed the field staff more time to engage with participants and the intervention components. Researchers were able to spend more time analyzing and preparing data for timely dissemination and had accurate, updated counts of study participants. It was also easier to identify subgroups for supplementary projects. For example, in the final year of the NSFS study, households with obese children were prioritized to receive a report card with BMI history.

This Web-based management solution can be used and tailored to a number of settings. It was successfully implemented with a group of data collectors, field staff, and researchers that had varying levels of literacy, English proficiency, and exposure to technology. All stakeholders reported that the application was easy to use and had a quick learning curve. Furthermore, this can be a low-cost data management solution, particularly for university-community partnerships. The application was developed by a computer science graduate student who joined the NSFS study as a research assistant and also provided back-end data management support. This approach is not without limitations. Foremost, the application requires access to the internet. When working in more remote parts of the community, data collectors reported that internet connections were not reliable. In these instances, data collectors would rely on printouts from the application as a backup. The application also required input from the computer programmer over time to add password-protected users and resolve minor issues. Finally, a limitation of the evaluation methodology was that this study was not able to completely differentiate between improvements that occurred because of data collectors gaining experience and application introduction. However, if the learning process for data collectors was ongoing, we would expect to see continued improvements in time efficiency after year 3 of data collection. The application introduction was a discrete change after year 2 and was not updated over time.

Conclusions

A Web-based management application was successful in improving data collection time efficiency and engagement among data collectors. This management solution was easily used by a varied audience and can be adapted to support a number of settings. As previously noted, “In an era of severe funding constraints for public health research, more efficient means of conducting research will be needed if scientific progress is to continue” [13].

Acknowledgments

The authors would like to acknowledge the contributions of the community-based data collectors and field staff for their role in developing the application described in this study. This material is based upon work that is supported by the National Institute of Food and Agriculture, US Department of Agriculture, under award number 2011-68001-30167. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the view of the US Department of Agriculture.

Conflicts of Interest

None declared.

References

1. Wallerstein NB, Duran B. Using community-based participatory research to address health disparities. Health Promot Pract 2006 Jul;7(3):312-323. [doi: 10.1177/1524839906289376] [Medline: 16760238]
2. Cacari-Stone L, Wallerstein N, Garcia AP, Minkler M. The promise of community-based participatory research for health equity: a conceptual model for bridging evidence with policy. Am J Public Health 2014 Sep;104(9):1615-1623. [doi: 10.2105/AJPH.2014.301961] [Medline: 25033119]
3. Lichtveld M, Kennedy S, Krouse RZ, Grimsley F, El-Dahr J, Bordelon K, et al. From design to dissemination: implementing community-based participatory research in postdisaster communities. Am J Public Health 2016 Jul;106(7):1235-1242. [doi: 10.2105/AJPH.2016.303169] [Medline: 27196662]

4. Israel BA, Eng E, Schulz AJ, Parker EA. Methods in Community-Based Participatory Research for Health. San Francisco: Jossey-Bass; 2005.
5. Jessell L, Smith V, Jemal A, Windsor L. Using technology to facilitate collaboration in community-based participatory research (CBPR). J Technol Hum Serv 2016;34(3):241-255 [FREE Full text] [doi: 10.1080/15228835.2016.1186581] [Medline: 29056871]
6. Zhang J, Sun L, Liu Y, Wang H, Sun N, Zhang P. Mobile device-based electronic data capture system used in a clinical randomized controlled trial: advantages and challenges. J Med Internet Res 2017 Mar 08;19(3):e66 [FREE Full text] [doi: 10.2196/jmir.6978] [Medline: 28274907]
7. Vallières F, McAuliffe E, van Bavel B, Wall PJ, Trye A. There's no app for that: assessing the impact of mHealth on the supervision, motivation, engagement, and satisfaction of community health workers in Sierra Leone. Ann Glob Health 2016;82(5):936-949. [doi: 10.1016/j.aogh.2016.07.002] [Medline: 28283148]
8. de la Torre A, Sadeghi B, Green RD, Kaiser LL, Flores YG, Jackson CF, et al. Niños Sanos, Familia Sana: Mexican immigrant study protocol for a multifaceted CBPR intervention to combat childhood obesity in two rural California towns. BMC Public Health 2013 Oct 31;13:1033 [FREE Full text] [doi: 10.1186/1471-2458-13-1033] [Medline: 24172250]
9. Bers TH. The popularity and problems of focus group research. Coll Univ 1989;64:260-268.
10. Saumure K. Focus groups: An overview. School of Library and Information Studies, University of Alberta 2001:2001.
11. Strauss AL. Qualitative Analysis for Social Scientists. Cambridge: Cambridge University Press; 1987.
12. Manzo RD, Rangel MI, Flores YG, de la Torre A. A community cultural wealth model to train promotoras as data collectors. Health Promot Pract 2018 May;19(3):341-348. [doi: 10.1177/1524839917703980] [Medline: 28385081]
13. Muennig PA. How automation can help alleviate the budget crunch in public health research. Am J Public Health 2015 Sep;105(9):e19-e22. [doi: 10.2105/AJPH.2015.302782] [Medline: 26180952]

Abbreviations

BMI: body mass index
NSFS: Niños Sanos, Familia Sana
PAR: participatory action research
UC: University of California
UCCE: UC Cooperative Extension
WHCC: West Hills Community College

Edited by G Eysenbach; submitted 25.04.18; peer-reviewed by J Zhang, S Tao; comments to author 09.06.18; revised version received 19.10.18; accepted 05.11.18; published 16.01.19.

Please cite as:
Hanbury MM, Sadeghi B, Tseregounis IE, Gomez-Camacho R, Manzo RD, Rangel MI, Alexandrescu B, de la Torre A
A Web-Based Application to Improve Data Collection in an Interventional Study Targeting Childhood Obesity: Pre-Post Analysis
J Med Internet Res 2019;21(1):e10861
URL: http://www.jmir.org/2019/1/e10861/
doi:10.2196/10861
PMID:30664465

Original Paper

Digital Exclusion Among Mental Health Service Users: Qualitative Investigation

Ben Greer[1,2], BSc (Hons); Dan Robotham[1,3], MSc, PhD; Sara Simblett[1,2], PhD, DClinPsy; Hannah Curtis[1,2], BSc (Hons); Helena Griffiths[1,2], MSc; Til Wykes[1,2], DBE, MPhil, DPhil

[1]Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[2]National Institute for Health Research Maudsley Biomedical Research Centre, South London and Maudsley National Health Service Foundation Trust, London, United Kingdom

[3]The McPin Foundation, London, United Kingdom

Corresponding Author:
Ben Greer, BSc (Hons)
Department of Psychology
Institute of Psychiatry, Psychology and Neuroscience
King's College London
16 De Crespigny Park
London,
United Kingdom
Phone: 44 0207848 ext 0423
Email: ben.greer@kcl.ac.uk

Abstract

Background: Access to internet-enabled technology and Web-based services has grown exponentially in recent decades. This growth potentially excludes some communities and individuals with mental health difficulties, who face a heightened risk of digital exclusion. However, it is unclear what factors may contribute to digital exclusion in this population.

Objective: To explore in detail the problems of digital exclusion in mental health service users and potential facilitators to overcome them.

Methods: We conducted semistructured interviews with 20 mental health service users who were deemed digitally excluded. We recruited the participants from a large secondary mental health provider in South London, United Kingdom. We employed thematic analysis to identify themes and subthemes relating to historical and extant reasons for digital exclusion and methods of overcoming it.

Results: There were three major themes that appeared to maintain digital exclusion: a perceived lack of knowledge, being unable to access the necessary technology and services owing to personal circumstances, and the barriers presented by mental health difficulties. Specific facilitators for overcoming digital exclusion included intrinsic motivation and a personalized learning format that reflects the individual's unique needs and preferences.

Conclusions: Multiple factors contribute to digital exclusion among mental health service users, including material deprivation and mental health difficulties. This means that efforts to overcome digital exclusion must address the multiple deprivations individuals may face in the offline world in addition to their individual mental health needs. Additional facilitators include fostering an intrinsic motivation to overcome digital exclusion and providing a personalized learning format tailored to the individual's knowledge gaps and preferred learning style.

(J Med Internet Res 2019;21(1):e11696) doi:10.2196/11696

KEYWORDS

digital exclusion; digital divide; digital inequality; technology; eHealth; mental health; social exclusion; mobile phone

Introduction

Internet use is near-ubiquitous in the United Kingdom with 80% of adults reporting using the internet daily [1]. The number of UK adults reporting having never accessed the internet is also declining, from 11.4% and 10.2% in 2015 and 2016, respectively [2,3], to 9% in 2017 [4]. Despite these decreases, a minority of individuals remain digitally excluded. There are inconsistent

definitions of digital exclusion [5], but it may be broadly defined as being unable to access or use internet-enabled technology and Web-based services [6].

Digital exclusion has been associated with material deprivation in the offline world. For example, Helsper [7] theorized that material deprivation in areas such as economic capital and sociocultural affiliations, mediated by individual factors such as access and skills, influences digital exclusion. In support of this model, Longley and Singleton [8] mapped material deprivation [9] onto geographical areas in England characterized by low levels of digital engagement and found that high material deprivation areas had lower levels of engagement in internet-enabled technology. Evidence from a recent international survey in Sweden and Britain also indicates that digital exclusion is becoming increasingly concentrated among vulnerable populations, including those who are socially isolated and unemployed [10]. In addition, factors such as old age [11,12] and living in rural areas [13,14] have previously been associated with digital exclusion.

Digital technology and services, including electronic patient record systems, self-monitoring technology, and Web-based therapies [15], have been proposed as pivotal components of planned improvements to mental health provision in the United Kingdom over the next 5 years [16]. These technologies and services have the potential to improve communication between service users and health care professionals, empower service users to more actively manage their health, and augment clinical decision making with real-time clinical information [17]. For these benefits to be realized, mental health service users must be able to access and use these systems.

Surveys conducted in the United Kingdom and the United States indicate that digital exclusion is more prevalent among those with mental health difficulties compared with those without [18,19]. Globally, numerous social determinants of health are associated with poorer mental health, including indicators of lower socioeconomic status such as poverty and employment status [20,21]. Consistent with theories of social and digital exclusion [7], people with mental health difficulties may be more likely to experience digital exclusion because they are also more likely to be socially excluded [22,23]. Evidence also suggests that despite technological developments [24], the factors associated with digital exclusion have been consistent over time. For example, in a sample of outpatients diagnosed with schizophrenia, depression, or anxiety disorders, only 36% reported having ever used the internet, with the primary barriers to internet use including financial costs, lack of skills or knowledge, cognitive difficulties, and access [25]. Ennis et al [26], using a survey of community mental health service users, reported that factors related to knowledge and access were associated with digital exclusion. In a 5-year follow-up of that study, Robotham et al [6] reported that while self-reported rates of digital exclusion had declined, barriers to inclusion still existed. Consistent with Borzekowski et al [25] and Ennis et al [26], key barriers included a lack of knowledge, skills, and financial resources.

To ensure all mental health service users can reap the benefits of digital services for health care and in daily life, it is important to understand the factors that lead to the exclusion of some, as this may help identify specific targets for intervention. Due to the pace of development in internet-enabled technology and Web-based services, it is important to gain a contemporary understanding of why some individuals remain excluded. Qualitative approaches may offer a greater depth of understanding of how the barriers discovered in the surveys emerge and interact to cause digital exclusion. The aim of this study was to explore reasons for digital exclusion among mental health service users and potential facilitators to overcome it.

Methods

Design

This was an exploratory qualitative study. We employed semistructured interviews using a topic guide. Service users and carers who required assistance with accessing and using internet-enabled technology and services attended a computer skills session and reviewed all study materials and measures. Members of the research team facilitated these sessions. We obtained ethical approval from the North West-Haydock Research Ethics Committee (16/NW/0792). The Feasibility and Acceptability Support Team for Researchers, a team of mental health service users and carers who have been specially trained to advise on research proposals and documentation, reviewed the study protocol, information sheet, and consent form for readability and acceptability.

Participants

We recruited participants from a large secondary mental health provider in South London, covering a diverse geographical area including areas of high poverty and urban deprivation. The participant recruitment sites included outpatient services, a community center, and weekly computer skills workshops where we recruited 4, 11, and 5 patients, respectively. We conducted recruitment and analysis concurrently. During recruitment, four authors (BG, DR, SS, and HC) met to discuss emerging themes and participant characteristics and, subsequently, employed maximum variation sampling [27] to recruit individuals who differed from current participants in terms of demographics and clinical diagnoses. We stopped recruitment when data saturation, the point at which interviews stopped yielding new themes, appeared to have been reached [28].

The inclusion criteria were being a current user of mental health services, possessing the capacity to provide informed consent (determined by the researcher trained in taking informed consent), and being digitally excluded (determined through a screening questionnaire, see below). The exclusion criteria were having a diagnosis of dementia, which is an illness that has specific needs in relation to digital intervention [29], and being under the age of 18, as previous research suggests true digital exclusion is less likely in this age group [30].

Screening Questionnaire

The screening questionnaire assessed potential participants' access and confidence with internet-enabled technology to determine whether they were digitally excluded (see Table 1). There was one question asking participants whether anything stopped them from using the internet. We considered participants

who answered yes to at least one of these items as digitally excluded. Potential participants who did not indicate any barriers or difficulties using the internet were considered ineligible to participate.

Interview Topic Guide

We based our interview topic guide (see Multimedia Appendix 1) on Robotham et al [6]. This began with an initial exploration of participants' familiarity with the internet (eg, "Have you ever used the internet?" and "In your own words can you describe what the internet is?"). This was a warm-up discussion for the remainder of the interview and provided the interviewer with an overall understanding of the participant's level of digital exclusion. We followed this with a discussion of potential barriers they may have encountered, as identified by Robotham et al [6] (eg, financial, knowledge, or skills), and facilitators to overcome digital exclusion (eg, delivering support in a group format vs one-on-one).

Procedure

Following successful screening for digital exclusion, participants took part in interviews conducted in a location of their choosing, including community centers (n=11), hospitals (n=5), and private interview rooms (n=4). The interviews were audio recorded, and if the participant did not provide consent for audio recording (n=1), the researcher transcribed the participant's responses in summary during the interview. The participants granted consent to use their electronic health records to obtain their demographic characteristics and diagnoses.

Thematic Analysis

We transcribed interviews verbatim with personally identifiable content omitted and used NVivo 11 software (QSR International, Melbourne, Australia) [31] for thematic analysis [32]. Initially, one author (BG) read and reread the transcripts, generating an initial list of codes based on the semantic content of the transcripts. The authors collated these into a list of candidate themes and subthemes through consultation with each other. Another author (HG) then independently double-coded the transcripts using the candidate themes as a framework. BG and HG discussed any discrepancies between their codes until they reached a consensus and revised the themes into their final format.

Table 1. Screening questionnaire assessing access and confidence with internet-enabled technology (N=20).

Question	Responses, n (%)
Which of the following items are you familiar with?	
Computer	3 (15)
Computer and smartphone	4 (20)
Computer, tablet, and smartphone	13 (65)
Which of the following items do you own?	
Computer	7 (35)
Computer and smartphone	4 (20)
Tablet and smartphone	1 (5)
None of them	8 (40)
Which of the following items do you have access to?	
Computer	13 (65)
Computer and smartphone	5 (25)
Computer, tablet, and smartphone	1 (5)
None of them	1 (5)
Is there anything that stops you from using the internet?	
Lack of knowledge about how to use technology	17 (85)
Not wanting to use technology	5 (25)
Lack of available technology	10 (50)
Lack of places to access technology	8 (40)
Fear of technology	10 (50)
Lack of credit or money	12 (60)
Security concerns	13 (65)
Other (participant-reported "confusion")	1 (5)

Results

Participants

We approached 36 individuals, of whom 20 provided informed consent and were interviewed. The final sample contained 13 men and 7 women with a mean age of 56.7 years (SD 11.3; range 39-80). Additional participant characteristics are shown in Table 2. On average, participants had been in contact with mental health services for 19 years.

Factors Maintaining Digital Exclusion

Knowledge

A perceived lack of knowledge was one of the most commonly reported barriers to engagement with internet-enabled technology. Participants reported confusion over how to use Web-based services, in addition to the internet-enabled technology itself:

> *What's the reason I haven't used the internet yet? I just can't get my head around it. I don't seem to be able to understand it so, I've put it on the back burner since I can't understand it, you know?* [P019, man, 57]

> *I just think God; how do they do that? You know, but where is it? How do you get it? How do you set it up? I suppose it is on the computer, is it? And you are on your phone, and all that, and I just think, you know. I don't know what half the things are, when they say "oh, do this, do that." I'm completely ignorant.* [P014, woman, 71]

Some participants also expressed uncertainty regarding potential sources of help for overcoming their digital exclusion, with those who did suggest sources of help (n=11) typically suggesting library services:

> *I don't know I don't know where to ask for help, I don't.* [P004, woman, 57]

> *Maybe the library, not sure? [laughter] perhaps the library?...Well I would ask if somebody can help me on the internet I suppose. Some of them are quite literate, computer literate and a lot more than I am so. But apart from that no I have no idea, but I mean if there are people I need to know [laughter].* [P010, woman, 52]

Personal Circumstances

Participants reported a range of personal circumstances as reasons for their digital exclusion. Perceived financial barriers were evident, including being unable to afford internet-enabled devices and accompanying services such as broadband:

> *Yeah if I could use a computer from home yeah, it's about affording it as well you know what I mean all the internet, broadband and stuff like that, yeah yeah.* [P007, man, 52]

> *I'm on benefits you know, that's it really…Yeah, once I have bills and all that and…event…and this that…I have nothing left for myself. And once you budget food out as well…you know what I mean?* [P020, man, 61]

Participants also reported barriers relating to their living situation, though it was unclear to what extent this may be related to financial costs:

> *Because I because I can't get...Wi-Fi...at home so I have to go to the library with my Chromebook.* [P004, woman, 57]

> *Well, because I live in a house which is split into studios, and they don't allow you to have an internet connection there you're not allowed to get Sky or broadband brought into the building.* [P017, man, 39]

Mental Health

Participants reported how their mental health difficulties, specifically psychosis, impacted on their digital exclusion. This included relapses and hallucinations preventing them from being able to use internet-enabled technology and forgetting how to use the technology. These memory difficulties also appeared to have hindered previous attempts to overcome digital exclusion:

> *I'm not sure whether I'd be able to use a computer if I had a relapse because the last time I was all over the place it was terrible.* [P002, man, 60]

> *Cos I'm schizophrenic you see so, some days I might be hearing voices and so I might not be able to or things like that, you know.* [P006, man, 59]

> *Because as I'm a psychotic patient you know I find it very difficult to remember things.* [P004, woman, 57]

> *No, because they used to—we used to—do a group here on computers and this is like 2012 and once a week, but every week I would like forget what I learnt cos we were only once a week, once a week is not good enough to do them sort of things you know?* [P019, man, 57]

In addition to the impact of mental health difficulties themselves, periods of time spent in inpatient care were also reported to be detrimental to participants' awareness of advances in technological development:

> *Well I think I think it's mainly the gaps, when you're mentally ill it's not like popping in to hospital for, cos I've had an operation as well.Wherethey're really keen to keep you, sometimes you're ill for months and you know you get gaps and stuff and as I say things just move on so fast. There was Twitter and things like that and apps I've no idea what all those things are.* [P012, woman, 57]

Table 2. Participant characteristics (N=20).

Characteristics	Participants, n (%)
Primary diagnosis	
Psychosis	10 (50)
Affective disorder	4 (20)
Personality disorder	1 (5)
Eating disorder	1 (5)
Other	4 (20)
Highest education achieved	
Secondary but not exam qualifications	4 (20)
Secondary (ordinary level or General Certificate of Secondary Education equivalent)	6 (30)
Secondary (Advanced level equivalent)	2 (10)
Vocational education or college	2 (10)
Higher-level qualification (eg, university degree or professional qualification)	6 (30)
Current occupation or employment	
Employed part-time	1 (5)
Unemployed	9 (45)
Student	2 (10)
Volunteering	5 (25)
Retired	3 (15)
Ethnicity	
White British	7 (35)
Black British	4 (20)
Asian	2 (10)
White other	2 (10)
Black other	4 (20)
Mixed	1 (5)

Overcoming Digital Exclusion

Motivation

There was variation in participants' motivation to overcome their digital exclusion. Some participants reported that digital exclusion did not negatively impact their lives and, therefore, did not express a desire to overcome it. For some who did report a desire to overcome their digital exclusion, it appeared that a perceived external pressure was an underlying factor:

> *It's not actually providing anything for me, it's just a machine it's not providing anything for me.* [P006, man, 59]

> *I find that otherorganizationstoo, banks and so on, more and more indicating that the preferred way for theircustomersto deal with them is through the internet...and that is a steady pressure as it were and is a growing pressure…and we may get to a stage where people not using the internet are such a minority that they're- they're disregarded.* [P015, man, 80]

Age also appeared to be a moderating factor in participants' motivation, with some older individuals being comfortable with digital exclusion. However, when asked whether efforts should be made to support older individuals in overcoming their digital exclusion, participants agreed that this population should not be overlooked:

> *Yes, I, I can easily say I would be open to that, whether I will actually grasp at it when the—when it was offered on the table in front of me…I'm not so sure, but it depends on what else was going on in my life* [P015, man, 80]

> *Yes, of course, yes absolutely. We're out there, you know, we're still alive and kicking, maybe sort of a bit slower—but we are out there.* [P016, woman, 75]

Personal Support Requirements

Participants expressed support for efforts to overcome digital exclusion that incorporated one-on-one and group support. Participants identified the social aspects of these approaches to support as key benefits:

> *Something positive to do somewhere positive to go. Don't have to be stuck indoors all the time I can get out and meet people and learn new skills if you like.* [P001, man, 47]

> *Those kind of you know you can do a course online, mental health management and all this stuff online, to me it's very isolating I prefer to be in a group and doing that kind of stuff rather than sitting at a computer doing it, so I do see a lot of the mental health kind of like workshops or courses or, you know, how to manage your mental health and all that and I just don't even bother looking at them, cos I just know I'm not gonna do it.* [P009, woman, 52]

However, participants also expressed some concerns regarding group support, including the level of individual support that could be offered in a group format compared with one-on-one:

> *Well you wouldn't be, the tutor wouldn't necessarily be able to give you his or her undivided attention, if there's a group.* [P016, woman, 75]

Where efforts to overcome digital exclusion may extend over multiple sessions, participants also voiced concerns regarding the level of commitment required and the potential consequences of missing a session:

> *I think for me in particular I think courses is something that you've got to kind of, you know you've got to commit to it I think for me it would be more like drop-in sessions on a particular subject. And more kind of one-to-one basis than than in a classroom…because especially with certain mental illnesses, mine in particular, I think if it's a class I become quite anxious if I'm not able to attend it. And then going back to it after not attending would create a lot of anxiety for me.* [P017, man, 39]

Discussion

Factors Maintaining Digital Exclusion

This study identified three major themes maintaining digital exclusion among mental health service users: knowledge, personal circumstances, and mental health. Similar themes have been reported in previous surveys of digitally excluded mental health service users [6,25]. By employing interviews, this study extends these earlier findings by identifying the specific barriers that exist within these overarching themes. Participants in this study also reported additional barriers, including relapses, memory difficulties, and periods of time spent in inpatient care, not previously identified in past research in relation to mental health.

Participants reported that both material deprivation (eg, personal finances and living situation) and aspects of their mental health were barriers to engagement with internet-enabled technology. This means that efforts to overcome digital exclusion among individuals with mental health difficulties should address both the multiple deprivations that digitally excluded individuals may experience in the offline world and their specific mental health needs. For example, hallucinations and poor memory may negatively impact on an individual's ability to engage and retain acquired skills in the future. So, understanding specific health needs may facilitate both short- and long-term digital inclusion. Gaps in knowledge and familiarity in people who have been isolated from developments in internet-enabled technology and services, such as those in inpatient services, should also be addressed. Providing supervised access to these technologies and services during periods of inpatient treatment may prevent these gaps arising, although this may not be required for everyone receiving inpatient mental health treatment. For example, people receiving treatment within acute mental health services, where the average length of stay is typically short, may experience a brief disruption in their access to internet-enabled technology but are unlikely to be completely excluded when returning to the community.

Overcoming Digital Exclusion

The range of reported knowledge gaps suggests that a single umbrella approach to overcoming digital exclusion may not be effective because it presupposes an equal level of understanding and confidence. While some may benefit from assistance tailored specifically toward internet use, others may require fundamental instruction in operating internet-enabled technology. Formally evaluating individuals' perceived confidence and competence would be beneficial to inform the scope and content of approaches to overcome digital exclusion. This would enable specific, individualized goals concerning technology and internet use to be set. Finding effective ways of ensuring digitally excluded individuals are aware of sources of support is also necessary.

Motivation was a key facilitator for overcoming digital exclusion but depended on a perceived disadvantage of being excluded. Without intrinsic motivation, engagement with programs to overcome digital exclusion will not be successful. For individuals expressing ambivalence about internet-enabled technology, techniques such as motivational interviewing [33,34] could be employed to foster this intrinsic motivation. There may be some instances, where digital exclusion does not exert a negative impact on the individual's life, when efforts to overcome digital exclusion are not necessary or desired by the individual. In addition, older people are often thought to be unmotivated to use internet-enabled technology [11], but the findings of this study indicate that this is not always the case.

Personal support was a facilitator, but participants did not all value the same format of support. Efforts to overcome digital exclusion would benefit from an individually tailored learning approach. There was a preference for individual support, but if a one-on-one approach is not feasible, then small groups could facilitate individual support more readily than large groups. In addition, the risk of individuals falling behind owing to an absence or inability to engage should also factor into this learning approach. This could be mitigated by supplementary written materials or the option to refresh their learning, an option endorsed by multiple participants.

Participants reported financial barriers but did not specify what they believed the cost of internet-enabled technology and services to be. This should be explored, as individuals may overestimate this cost compared with the cost of available technology and services. Where financial barriers do exist,

individuals could be signposted to free sources of internet-enabled technology and services, such as libraries and other public services. The option to supply these devices on a temporary or permanent basis could be explored but would need to be considered against the number of individuals requiring this support.

Limitations and Future Research

Previous research highlights that digital inclusion is not always permanent, with some individuals who were previously digitally engaged subsequently disengaging [35]. There were 14 participants in this study who reported previously using the internet; however, this study did not compare whether the reasons for disengagement differ from those of never engaging with the internet or whether the consequences of each type of exclusion differ. Future research should, therefore, ensure that participants' status as digitally naïve or digitally disengaged is identified, as there may be shared and divergent causes and consequences of both types of exclusion.

Reported difficulties with hallucinations are arguably specific to those with a diagnosis of psychosis, but other difficulties, such as amotivation and anhedonia, are shared across multiple diagnoses and may also impact on the ability to engage with internet-enabled technology and services. As our sample had 50% (10/20) participants with a psychosis diagnosis, this allowed both these specific and shared symptoms to be reported as affecting digital exclusion. Future research could benefit from a more comprehensive formulation of individuals' difficulties beyond the level of diagnosis, to identify and explore the impact of specific clinical difficulties on digital exclusion.

We recruited 25% (5/20) participants from computer skills workshops, and they may have possessed a higher level of intrinsic motivation and knowledge than the other participants. This group allowed the exploration of these different potential motivational effects, and shared factors were identified between this group and the remaining group who had not experienced such training.

Future research should consider whether there are any potential variables, such as demographic characteristics, that may relate to digital exclusion. Exploring how these factors have historically impacted on digital exclusion could help to identify at-risk groups and inform personally tailored approaches to overcoming digital exclusion.

Conclusion

This study identified three themes that maintain digital exclusion among mental health service users: knowledge, personal circumstances, and mental health. Efforts to support mental health service users to overcome digital exclusion must, therefore, address the multiple deprivations individuals may face in the offline world and their individual mental health needs. In addition to addressing wider societal issues related to finance and living circumstances, specific facilitators for overcoming digital exclusion include fostering an intrinsic motivation to overcome it and a personalized learning format tailored to individuals' specific knowledge gaps and preferred learning style.

Acknowledgments

This paper represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley National Health Service Foundation Trust and King's College London. TW would also like to acknowledge the support of her NIHR Senior Investigator award. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Topic guide used in participant interviews.

[PDF File (Adobe PDF File), 13KB - jmir_v21i1e11696_app1.pdf]

References

1. Office for National Statistics. Internet users in the UK: 2017. London: Office for National Statistics; 2017 May 19. URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2017 [WebCite Cache ID 75IUNR6hJ]
2. Office for National Statistics. Internet users: 2015. London: Office for National Statistics; 2015 May 22. URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2015 [WebCite Cache ID 75IURPUZa]
3. Office for National Statistics. Internet users in the UK: 2016. London: Office for National Statistics; 2016 May 20. URL: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/bulletins/internetusers/2016 [WebCite Cache ID 75IUUklr4]
4. Office for National Statistics. Internet access-households and individuals, Great Britain: 2017. London: Office for National Statistics; 2017 Aug 03. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2017 [WebCite Cache ID 75IUY9clg]

5. Scheerder A, van Deursen A, van Dijk J. Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telematics and Informatics 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
6. Robotham D, Satkunanathan S, Doughty L, Wykes T. Do We Still Have a Digital Divide in Mental Health? A Five-Year Survey Follow-up. J Med Internet Res 2016 Dec 22;18(11):e309 [FREE Full text] [doi: 10.2196/jmir.6511] [Medline: 27876684]
7. Helsper EJ. A Corresponding Fields Model for the Links Between Social and Digital Exclusion. Commun Theor 2012 Oct 15;22(4):403-426. [doi: 10.1111/j.1468-2885.2012.01416.x]
8. Longley PA, Singleton AD. Linking Social Deprivation and Digital Exclusion in England. Urban Studies 2009 May 06;46(7):1275-1298. [doi: 10.1177/0042098009104566]
9. Ministry of Housing, Communities and Local Government. Index of Multiple Deprivation 2004. London: Ministry of Housing, Communities and Local Government; 2004 Jan 06. URL: https://data.gov.uk/dataset/59599787-bd50-4500-a409-fc586260dbbd/index-of-multiple-deprivation-2004 [WebCite Cache ID 75IUegUPa]
10. Helsper EJ, Reisdorf BC. The emergence of a “digital underclass” in Great Britain and Sweden: Changing reasons for digital exclusion. New Media & Society 2016 Mar 03;19(8):1253-1270. [doi: 10.1177/1461444816634676]
11. Olson KE, O'Brien MA, Rogers WA, Charness N. Diffusion of Technology: Frequency of Use for Younger and Older Adults. Ageing Int 2011 Mar;36(1):123-145 [FREE Full text] [doi: 10.1007/s12126-010-9077-9] [Medline: 22685360]
12. van Deursen AJ, Helsper EJ. A nuanced understanding of Internet use and non-use among the elderly. European Journal of Communication 2015 Apr 09;30(2):171-187. [doi: 10.1177/0267323115578059]
13. Philip L, Cottrill C, Farrington J, Williams F, Ashmore F. The digital divide: Patterns, policy and scenarios for connecting the ‘final few’ in rural communities across Great Britain. Journal of Rural Studies 2017 Aug;54:386-398. [doi: 10.1016/J.JRURSTUD.2016.12.002]
14. Williams F, Philip L, Farrington J, Fairhurst G. 'Digital by Default and the 'hard to reach: Exploring solutions to digital exclusion in remote rural areas. Local Economy 2016 Sep 30;31(7):757-777. [doi: 10.1177/0269094216670938]
15. Wykes T, Brown M. Over promised, over-sold and underperforming? - e-health in mental health. J Ment Health 2016;25(1):1-4 [FREE Full text] [doi: 10.3109/09638237.2015.1124406] [Medline: 26732733]
16. NHS England. The Five Year Forward View For Mental Health. London: Mental Health Task Force; 2016 Feb. URL: https://www.england.nhs.uk/wp-content/uploads/2016/02/Mental-Health-Taskforce-FYFV-final.pdf [WebCite Cache ID 75IUjXOy6]
17. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
18. Dobransky K, Hargittai E. Unrealized potential: Exploring the digital disability divide. Poetics 2016 Oct;58:18-28. [doi: 10.1016/J.POETIC.2016.08.003]
19. Tobitt S, Percival R. Switched on or switched off? A survey of mobile, computer and Internet use in a community mental health rehabilitation sample. J Ment Health 2017 Jul 04:1-7. [doi: 10.1080/09638237.2017.1340623] [Medline: 28675329]
20. Lund C, Brooke-Sumner C, Baingana F, Baron EC, Breuer E, Chandra P, et al. Social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews. Lancet Psychiatry 2018 Apr;5(4):357-369. [doi: 10.1016/S2215-0366(18)30060-9] [Medline: 29580610]
21. Molarius A, Berglund K, Eriksson C, Eriksson HG, Lindén-Boström M, Nordström E, et al. Mental health symptoms in relation to socio-economic conditions and lifestyle factors--a population-based study in Sweden. BMC Public Health 2009 Aug 20;9:302 [FREE Full text] [doi: 10.1186/1471-2458-9-302] [Medline: 19695085]
22. Morgan C, Burns T, Fitzpatrick R, Pinfold V, Priebe S. Social exclusion and mental health: conceptual and methodological review. Br J Psychiatry 2007 Dec;191:477-483. [doi: 10.1192/bjp.bp.106.034942] [Medline: 18055950]
23. Pickett KE, Wilkinson RG. Income inequality and health: a causal review. Soc Sci Med 2015 Mar;128:316-326. [doi: 10.1016/j.socscimed.2014.12.031] [Medline: 25577953]
24. Government Office for Science. Technology and Innovation Futures 2017. London: Government Office for Science; 2017 Jan 23. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/584219/technology-innovation-futures-2017.pdf [WebCite Cache ID 75IUpXYlC]
25. Borzekowski DLG, Leith J, Medoff DR, Potts W, Dixon LB, Balis T, et al. Use of the internet and other media for health information among clinic outpatients with serious mental illness. Psychiatr Serv 2009 Sep;60(9):1265-1268. [doi: 10.1176/ps.2009.60.9.1265] [Medline: 19723745]
26. Ennis L, Rose D, Denis M, Pandit N, Wykes T. Can't surf, won't surf: the digital divide in mental health. J Ment Health 2012 Aug;21(4):395-403 [FREE Full text] [doi: 10.3109/09638237.2012.689437] [Medline: 22712756]
27. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
28. Fusch P, Ness L. Are we there yet? Data saturation in qualitative research. Qual Rep 2015 Sep 07;20(9):1408-1416 [FREE Full text]

29. Dementia and digital: Using technology to improve health and wellbeing for people with dementia and their carers. Sheffield: Tinder Foundation; 2016 Sep. URL: https://www.housinglin.org.uk/_assets/Resources/Housing/OtherOrganisation/dementia_and_digital.pdf [WebCite Cache ID 75IUyWuaD]
30. Hargittai E, Hinnant A. Digital Inequality: Differences in Young Adults' Use of the Internet. Communication Research 2008 Aug 04;35(5):602-621. [doi: 10.1177/0093650208321782]
31. QSR International. NVivo qualitative data analysis software.: QSR International Pty Ltd 2015 Version 11; 2015. URL: http://www.qsrinternational.com/nvivo/support-overview/faqs/how-do-i-cite-nvivo-for-mac,-nvivo-11-for-windows,
32. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Miller WR, Rollnick S. Motivational interviewing: Helping people change. 3rd edition. New York, NY: Guilford Press; Oct 05, 2012:1-482.
34. Rubak S, Sandbaek A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. Br J Gen Pract 2005 Apr;55(513):305-312 [FREE Full text] [Medline: 15826439]
35. Olphert W, Damodaran L. Older people and digital disengagement: a fourth digital divide? Gerontology 2013;59(6):564-570. [doi: 10.1159/000353630] [Medline: 23969758]

Edited by G Eysenbach; submitted 25.07.18; peer-reviewed by EB Davies, G Goggin; comments to author 05.09.18; revised version received 17.09.18; accepted 18.09.18; published 09.01.19.

Please cite as:
Greer B, Robotham D, Simblett S, Curtis H, Griffiths H, Wykes T
Digital Exclusion Among Mental Health Service Users: Qualitative Investigation
J Med Internet Res 2019;21(1):e11696
URL: http://www.jmir.org/2019/1/e11696/
doi:10.2196/11696
PMID:30626564

Original Paper

Differences Between Mothers and Fathers of Young Children in Their Use of the Internet to Support Healthy Family Lifestyle Behaviors: Cross-Sectional Study

Rachel Laws[1,2], BSc (Nutrition), MSc (Nutrition & Dietetics), PhD; Adam D Walsh[1], BAppSci (Human Movement), Master of Nutrition & Dietetics, PhD; Kylie D Hesketh[1,2], BBSc (Hons), PhD; Katherine L Downing[1], BHealthSc, BEx&SportSc (Hons), PhD; Konsita Kuswara[1], BMedSci, MSc (Nutrition & Dietetics), MPH; Karen J Campbell[1,2], BSc, MPH, PhD

[1]Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Geelong, Australia

[2]Centre for Research Excellence in the Early Prevention of Obesity in Childhood, Sydney, Australia

Corresponding Author:
Rachel Laws, BSc (Nutrition), MSc (Nutrition & Dietetics), PhD
Institute for Physical Activity and Nutrition
School of Exercise and Nutrition Sciences
Deakin University
221 Burwood Highway
Burwood, VIC
Geelong, 3125
Australia
Phone: 61 432388593
Fax: 61 392446017
Email: r.laws@deakin.edu.au

Abstract

Background: In early life, both mothers and fathers are important influences on their children's diet, active play, and obesity risk. Parents are increasingly relying on the internet and social media as a source of information on all aspects of parenting. However, little is known about the use of Web-based sources of information relevant to family lifestyle behaviors and, in particular, differences between mothers' and fathers' use and sociodemographic predictors.

Objective: The objective of this study was to examine if mothers and fathers differ in their use of the internet for information on their own health and their child's health, feeding, and playing and to examine sociodemographic predictors of the use of the internet for information on these topics.

Methods: We conducted a secondary analysis on data collected from mothers (n=297) and fathers (n=207) participating in the extended Infant Feeding, Activity and Nutrition Trial (InFANT Extend) when their children were 36 months of age. The main outcome variables were the use of the internet for information gathering for parents' own health and child health, feeding, and playing. Binary logistic regression was used to examine the sociodemographic predictors of outcomes.

Results: Compared with fathers (n=296), a higher proportion of mothers (n=198) used the internet for information on their own health (230, 78.5% vs 93, 46.5%), child health (226, 77.1% vs 84, 42.4%), child feeding (136, 46.3% vs 35, 17.5%), and child play (123, 42.1% vs 28, 14.0%) and intended to use Facebook to connect with other parents (200, 74.9% vs 43, 30.5%). Despite the high use of the internet to support family health behaviors, only 15.9% (47/296) of mothers reported consulting health practitioners for advice and help for their own or their child's weight, diet, or physical activity. Sociodemographic predictors of internet use differed between mothers and fathers and explained only a small proportion of the variance in internet use to support healthy family lifestyle behaviors.

Conclusions: Our findings support the use of the internet and Facebook as an important potential avenue for reaching mothers with information relevant to their own health, child health, child diet, and active play. However, further research is required to understand the best avenues for engaging fathers with information on healthy family lifestyle behaviors to support this important role in their child's life.

Trial Registration: ISRCTN Registry ISRCTN81847050; http://www.isrctn.com/ISRCTN81847050

(J Med Internet Res 2019;21(1):e11454) doi:10.2196/11454

KEYWORDS

child; family; healthy lifestyle; infant; internet; obesity; parents

Introduction

The prevention of childhood obesity remains a significant global public health challenge, with an estimated 41 million children under 5 years of age being overweight or obese [1]. In Australia, nearly one-quarter of 2-4 year-olds are overweight or obese [2], with important health and economic consequences [3] underscoring the importance of early prevention. Lifestyle risk factors for overweight and obesity in early life include a short breastfeeding duration [4], early introduction of solids [4], poor child diet quality [5], short sleep duration [5] as well as low levels of child physical activity [5] and high levels of sedentary behaviors [5], such as screen time exposure. Parents play a primary role in shaping these behaviors through parental modeling, parenting styles, and food, physical activity, and sedentary environments provided [6].

Although much of the existing research has focused on the maternal influences on children's lifestyle behaviors [7], recent research has also found fathers to have an important influence. Associations between the dietary intakes of fathers and their children have been identified from as early as 20 months of age [8], and these continue through early childhood [9]. Furthermore, parenting styles of fathers [10] as well as feeding practices [11] have been found to be associated with child adiposity [10]. This highlights the importance of considering both maternal and paternal influences on child obesity risk behaviors.

Providing parents with high-quality information and support to inform healthy family lifestyle behaviors is critical. Evidence suggests that parents are actively seeking support in their parenting role. On average, parents in Australia make 11 visits to general practitioners and 14 visits to Maternal and Child Health nurses in the first year of their child's life, with the majority of these visits unrelated to illness [12]. In addition to health professionals, parents are increasingly relying on the internet and social media for information and support on all aspects of parenting, including family health [13]. A Google survey in 2014 revealed that new and expecting parents undertake nearly 3 times the number of Web-based searches as nonparents [14]. In the United States, 3-quarters of parents use social media, both for information and social support [15]. In line with this, a nationally representative survey conducted in Australia in 2016 revealed that more than 60% of parents and 73% of those with children under 5 years of age used websites, blogs, and Web-based forums to get child health information in the last 6 months. Interestingly, 30% reported that they did not trust this information source [16].

Research on the use of the internet to support lifestyle behaviors important for child obesity prevention is growing but still limited. The use of the internet as the main source of information on nutrition in the general population has grown rapidly in the last decade [17], and research supports that this is also the case among parents of young children [18,19], including those at risk of obesity [20]. There is much less information on the use of the internet by parents to seek information on active childhood play and limiting sedentary behavior. Despite this, excessive screen time and insufficient physical activity rated the top and third top concern of Australian parents in a national health poll [21]. Just 1 Spanish study addresses this, showing that around 30% of parents of young children search for Web-based information on play activities [22]. Our own qualitative research [23] suggests that parents would welcome Web-based sources of support in this area.

Despite the important influence of fathers on child health behaviors [9,24,25], research on the use of the internet and social media to support healthy family lifestyle behaviors has almost exclusively been with mothers, with very few studies [22,26] focusing on fathers. Research suggests that fathers engage in Web-based activities significantly more for general rather than for parenting purposes [27]. In line with this, 1 study of parents of young children in Spain [22] has found that mothers were more likely to search for digital information on child development and family health compared to fathers, but there were no differences in seeking information on play activities. Another study [26] found no significant difference in the use of the internet for health information between expectant mothers and fathers; however, the sample of fathers was small (n=21), limiting the ability to draw robust conclusions. In terms of social media, Facebook was found to be the most popular among parents in the United States; however, it was used more by mothers (81%) compared with fathers (66%) for information and social support [15]. Further research is required to examine fathers' use of the internet to support healthy family lifestyle behaviors.

We also know little about how the use of the internet for family health information seeking varies according to parents' socioeconomic position, ethnicity, or age. There is a debate in the general parenting literature about the existence of a digital divide in the use of the internet and Web-based sources of health information according to parents' sociodemographic characteristics. Although internet access is near universal, some research suggests that parents of lower socioeconomic position (as indicated by education, income, or the use of services for vulnerable parents) are less likely to use the internet for health information seeking [22,28,29] but equally likely to use social media for these purposes [28]. In line with this, Guerra-Reyes and colleagues [30] reported that in postpartum women, college graduates searched authoritative Web-based sources, while nongraduates preferred forums. Other studies, however, have found no difference in the use of the internet for information on pregnancy [26] and personal or child health information according to parents' education levels [20]. Again, these studies were predominantly among mothers, with very few studies specifically examining sociodemographic predictors of fathers' use of the internet for family health information seeking.

The aim of this paper is to examine if mothers and fathers differ in the use of the internet for information on their own health and their child's health, feeding, and playing and to examine sociodemographic predictors of the use of the internet for information on these topics. This will provide new insights into the differing patterns of internet use by mothers and fathers to support healthy family lifestyle behaviors, which is important in informing the future development of targeted interventions in this area.

Methods

Study Context and Participant Recruitment

This study undertook a secondary analysis of data collected as part of the extended Infant Feeding, Activity and Nutrition Trial (InFANT Extend) Program: a cluster randomized controlled trial of a 33-month parent-focused child obesity prevention intervention. Details of the trial and participant recruitment have been reported elsewhere [31]. In brief, the trial aimed to test the effectiveness of a 6-session group-based program delivered to first-time parents when infants were 3-18 months of age, followed by quarterly newsletters from 18-36 months of age. Control participants received usual care. Participants were recruited from first-time parent groups within 7 purposively selected socioeconomically disadvantaged local government areas in Melbourne, Australia. Individual parents were eligible to participate if they gave informed written consent, were first to time parents, and were literate in English. Both the main caregiver and partner were invited to participate in the program and data collection. Infants with chronic health problems likely to influence height, weight, levels of physical activity, or eating habits were excluded from analyses but were permitted to participate in the program.

Data Collection

This study analyzed data collected from main caregivers and partners participating in the InFANT Extend trial when children were 36 months of age (between 2014-2015). Participants were asked about their use of the internet in general and for information on their own health, dieting, recipes, child health, how to feed their baby, and how to play with their baby. Participants were asked if and how often they used Facebook and whether they could imagine using Facebook to connect with other parents. Sociodemographic variables collected included parent education level, country of birth, relationship status, employment status, income, and self-rated health status. Participants' perception of the attention they paid to personal health habits and weight control were collected using a 5 point Likert scale from "none" to "very much." Participants were asked about how well they were coping with life at present from "not at all" to "extremely well." The main caregiver was also asked about whether they had consulted with a health practitioner for advice and help for their own or child weight, diet, or physical activity; the type of practitioner consulted as well as whether they had access to the internet, a tablet, desktop computer, laptop, or smartphone at home. All data were collected using paper-based surveys mailed to participants with paid reply envelopes.

Data Analyses

The main outcome variables of interest were the use of the internet for information on parents' own health and child health, feeding, and playing. Preliminary analyses showed no significant differences in these variables between the intervention and control group participants; hence, data were analyzed for both groups combined. However, the treatment group was controlled for in the analysis; main caregiver and partner data were analyzed separately. All main caregiver surveys were completed by mothers, and all partner surveys were completed by fathers with the exception of 5 surveys (2 mothers, 2 stepfathers, and 1 participant who did not identify their relationship with the child), which were excluded for the purposes of this analysis. Initial analyses were conducted to describe each sample (mothers and fathers) and their use of the internet. Binary logistic regression analyses were used to examine sociodemographic predictors of (1) mothers' use of the internet for information on child feeding and playing and (2) fathers' use of the internet for information on their own health and their child's health. These outcome variables were chosen for logistic regression analysis because they showed the most variability (around 40%-50% of the sample responding positively), whereas other variables had very high positive responses (>75%; ie, mothers' use of the internet for own health and child) or very low positive responses (<20%; ie, fathers' use of the internet for information on child feeding and playing), limiting the ability to examine predictors. The selection of sociodemographic and other independent variables to be included in the models was based on existing literature and hypotheses about factors potentially important in influencing internet use. Independent variables showing little variation (ie, relationship status and employment status in fathers showed >90% were married and worked full-time) were considered redundant and not included in the models.

Ethics

This study was approved by the Deakin University Human Ethics Research Committee.

Results

Participant Characteristics

A total of 57.8% (297/514) main caregivers and 47.8% (207/433) partners participating in the trial completed the survey when their child was 36 months of age. There were no significant differences in any baseline characteristics of those who completed the survey at 36 months and those lost to follow-up, with the exception that those lost to follow-up had a higher prepregnancy body mass index (BMI). Participant characteristics are shown in Table 1. A higher proportion of mothers had a university education than fathers. Just under half of mothers were not in paid employment and were looking after children full-time, and a similar proportion worked part-time. Just over 10% (31/297, 10.4%) of mothers worked full-time compared with over 90% (179/197, 90.9%) of fathers. Most parents were married or in a de facto relationship and around 80% (387/487, 79.5%) of all parents were born in Australia. Self-rated health status was high for both mothers and fathers,

and just under half rated their attention to personal health habits as high.

Table 1. Participant characteristics when their child was aged 3 years.

Characteristics	Mothers (n=297)	Fathers (n=202)
Age in years, mean (SD); range	35.2 (4.2); 22.3-47.9	37.3 (5.4); 26.2-57.8
Education, n (%[a])		
High school education or less	36 (12.4)	33 (16.8)
Trade, certificate, or diploma	80 (27.6)	78 (39.8)
University degree or higher degree	174 (60.0)	85 (43.4)
Relationship status, n (%[a])		
Married or de facto relationship	274 (92.3)	N/A[b]
Separated, divorced, or widowed	12 (4.0)	N/A
Never married	11 (3.7)	N/A
Employment status, n (%[a])		
Working full-time	31 (10.4)	179 (90.9)
Working part-time	132 (44.4)	9 (4.6)
Unemployed or laid off	1 (0.3)	4 (2.0)
Keeping house full-time	130(43.8)	2 (1.0)
Studying full-time	3 (1.0)	3 (1.5)
Health care card, n (%[a])	35 (11.8)	N/A
Country of birth, n (%[a])		
Australia	229 (78.7)	158 (80.6)
Other	62 (21.3)	38 (19.4)
Self-rated health status, n (%[a])		
Excellent	19 (6.4)	20 (9.9)
Very good or good	226 (76.1)	156 (77.2)
Fair	49 (16.5)	24 (11.9)
Poor	3 (1.0)	2 (1.0)
Attention to personal health habits, n (%[a])		
None or little	40 (13.5)	32 (15.8)
Some	121 (40.7)	78 (38.6)
Much or very much	136 (45.8)	92 (45.5)
Attention paid to controlling weight, n (%[a])		
None or little	90 (30.3)	55(27.4)
Some	115 (38.7)	82 (40.8)
Much or very much	92 (31.0)	64 (31.8)
Coping with life at present, n (%[a])		
Not at all or little	19 (6.4)	10 (5.0)
Fairly well	135 (45.5)	87 (43.1)
Very or extremely well	143 (48.1)	105 (52.0)

[a]Percentages relate to samples with valid data for each variable; missing data were excluded.

[b]N/A: not applicable (as these questions were only asked of mothers).

Use of the Internet to Support Healthy Family Lifestyle Behaviors

Among the 297 mothers, 292 (98.3%) reported having internet access at home and 256 (86.1%) had a tablet, 292 (98.3%) a smartphone, 254 (85.5%) a laptop, and 149 (50.2%) a desktop computer. Although almost all mothers and fathers reported using the internet, there were distinct differences in use of the internet for health and lifestyle-related information between mothers and fathers (Table 2). Over 90% of mothers reported using the internet for recipes and around 3-quarters for information on their own health and their child's health compared with less than half of fathers using the internet for their own health or their child's health. Around half of the mothers reported using the internet for information on child feeding and a similar proportion for play compared with only 17.5% (35/200) and 14.0% (28/200) of fathers, respectively. Around 90% (266/297, 89.9%) of mothers and 70% (140/198, 69.7%) of fathers reported being a member of Facebook, with frequent use in both groups. However, less than a third of fathers reported that they could imagine using Facebook to connect with other parents, compared with nearly 3-quarters of mothers.

Health Professional Consultations for Diet, Activity, and Weight

In contrast to the high proportion of mothers reporting using the internet for information on dieting, child feeding, and playing, only 15.9% (47/296) reported seeking health professional advice and help for their own or their child's weight, diet, or physical activity (Table 3); this question was only asked of primary caregivers. General practitioners (family doctors) and pediatricians were the most common health professionals consulted.

Table 2. Internet and Facebook use by mothers and fathers.

Use	Mothers (n=297), n (%[a])	Fathers (n=202), n (%[a])
Use the internet	296 (99.7)	198 (98.5)
Use the internet to access information on		
Your health	230 (78.5)	93 (46.5)
Dieting or your diet	109 (37.2)	59 (29.6)
Fitness	N/A[b,c]	81 (40.5)
Child health	226 (77.1)	84 (42.4)
How to feed your child	136 (46.3)	35 (17.5)
How to play with your child	123 (42.1)	28 (14.0)
Recipes	274 (92.9)	N/A[d]
Member of Facebook	266 (89.9)	140 (69.7)
Frequency of Facebook use		
Once a week or less	24 (9.4)	18 (12.9)
A few times a week	17 (6.4)	17 (12.2)
Once a day	57 (21.3)	38 (27.3)
Several times a day	168 (62.9)	66 (47.5)
Could you imagine using Facebook to connect with other parents		
Yes	200 (74.9)	43 (30.5)
No	19 (7.1)	65 (46.1)
Maybe	48 (18.0)	33 (23.4)

[a]Percentages relate to samples with valid data for each variable; missing data were excluded.

[b]N/A: not applicable.

[c]These questions were only asked of fathers.

[d]These questions were only asked of mothers.

Table 3. Mothers' use of health practitioners for advice and help for own weight or child weight, diet, or physical activity (N=296).

Mothers' use of health practitioners	Mothers, n (%)
Any health professional advice since birth	47 (15.9)
Maternal and child health telephone line	12 (4.1)
Mother and baby center (day stay)	6 (2.0)
Mother and baby center (overnight stay)	3 (1.0)
Home visit or outreach nurse	5 (1.7)
General practitioner or family doctor	23 (7.8)
Pediatrician	19 (6.4)
Dietitian	14 (4.7)
Chiropractor, naturopath, or osteopath	12 (4.1)
Other health professional	12 (4.1)

Predictors of Mothers' Use of the Internet for Information on Child Feeding and Playing

Predictors of mothers' use of the internet for information on child feeding and playing are shown in Tables 4 and 5 respectively. Nonworking mothers were 1.7 times more likely to seek information from the internet on child feeding compared with mothers who were working or studying, with employment status the only significant predictor in the model. Mothers born outside Australia were also more likely to use the internet for information on child feeding compared with Australian-born mothers; however, this did not reach statistical significance. The model explained 37%-50% of the variance in mothers' use of the internet for information on child feeding. There were no significant sociodemographic predictors of internet use for information on child play; however, there was a trend for university-educated mothers to be more likely to seek information on child play compared with their less educated counterparts (Table 5). This model explained only 25%-33% of the variance in mothers' use of the internet for information on child play.

Table 4. Predictors of internet use by mothers for information on child feeding when their child was aged 3 years.

Predictor	β (SE β)	Odds ratio (95% CI)	*P* value
Age	0.01 (0.03)	1.01 (0.95-1.07)	.87
Education			.54
Nonuniversity educated	Reference	1.00	
University educated	−0.16 (0.26)	0.85 (0.51-1.43)	
Employment status			.02
Working or studying	Reference	1.00	
Not working (keeping house or unemployed)	0.56 (0.25)	1.76 (1.08-2.87)	
Country of birth			.07
Australia	Reference		
Other	0.56 (0.32)	1.76 (0.95-3.25)	
Child body mass index z-score age 3 years	0.02 (0.15)	1.02 (0.76-1.37)	.91
Attention paid to personal health habits			.23
None, little, or some	Reference		
Much or very much	0.30 (0.25)	1.35 (0.83-2.20)	
Treatment group			.15
Control	Reference		
Intervention	0.36 (0.25)	1.43 (0.88-2.31)	

Table 5. Predictors of internet use by mothers for information on child play when their child was aged 3 years.

Predictor	β (SE β)	Odds ratio (95% CI)	*P* value
Age	−0.03 (0.03)	0.97 (0.91-1.03)	.33
Education			.09
Nonuniversity educated	Reference	1.00	
University educated	0.45 (0.27)	1.57 (0.93-2.66)	
Employment status			.19
Working or studying	Reference	1.00	
Not working (keeping house or unemployed)	0.33 (0.25)	1.39 (0.85-2.27)	
Country of birth			.87
Australia	Reference	1.00	
Other	0.05 (0.31)	1.05 (0.57-1.94)	
Child body mass index *z*-score age 3 years	−0.09 (0.15)	0.92 (0.68-1.24)	.57
Attention paid to personal health habits			.54
None, little, or some	Reference		
Much or very much	0.16 (0.25)	1.17 (0.72-1.90)	
Treatment group			.30
Control	Reference		
Intervention	0.26 (0.25)	1.29 (0.80-2.11)	

Predictors of Fathers' Use of the Internet for Information on Their Own Health and Their Child's Health

Fathers' reported attention to personal health habits was the only significant predictor of the use of the internet for their own health. There was a trend for fathers with higher levels of education to be more likely to use the internet for their own health; however, this did not reach statistical significance (Table 6). This model only explained 8%-11% of the variance in fathers' use of the internet for information on their own health. A younger age, higher education level, and greater reported attention to personal health habits were all significant predictors of the use of the internet for information on child health (Table 7). Fathers born outside Australia were also more likely to use the internet for child health, but this did not reach statistical significance. This model explained 13%-18% of the variance in fathers' use of the internet for information on their child's health.

Table 6. Predictors of use of the internet by fathers for information on their own health when their child was aged 3 years.

Predictor	β (SE β)	Odds ratio (95% CI)	*P* value
Age	−0.04 (0.03)	0.96 (0.90-1.02)	.16
Education			.08
Nonuniversity educated	Reference	1.00	
University educated	0.56 (0.32)	1.74 (0.93-3.27)	
Country of birth			.41
Australia	Reference	1.00	
Other	0.33 (0.40)	1.39 (0.64-3.00)	
Attention paid to personal health habits			
None or little	Reference	1.00	.02
Some	−0.01 (4.63)	0.99 (0.40-2.45)	.98
Much or very much	0.88 (0.45)	2.42 (0.99-5.88)	.05[a]
Fathers' body mass index	0.04 (0.05)	1.04 (0.95-1.14)	.37
Treatment group			.61
Control	Reference	1.00	
Intervention	−0.16 (0.31)	0.85 (0.47-1.56)	

Table 7. Predictors of use of the internet by fathers for information on their child's health when their child was aged 3 years.

Predictor	β (SE β)	Odds ratio (95% CI)	*P* value
Age	−0.07 (0.03)	0.93 (0.87-0.99)	.03
Education			.001
Nonuniversity educated	Reference		
University educated	1.07 (0.34)	2.92 (1.51-5.66)	
Country of birth			.08
Australia	Reference		
Other	0.71 (0.41)	2.04 (0.91-4.55)	
Attention paid to personal health habits			
None or little	Reference		.05
Some	0.97 (0.53)	2.62 (0.93-7.41)	.07
Much or very much	1.26 (0.52)	3.52 (1.28-9.67)	.02
Fathers' body mass index	0.004 (0.05)	1.004 (0.91-1.11)	.93
Treatment group			.93
Control	Reference		
Intervention	−0.03 (0.32)	0.97 (0.52-1.83)	

Discussion

Principal Findings

This is 1 of few studies to compare mothers' and fathers' use of the internet to seek information on their own health and their child's health, feeding, and playing and to examine sociodemographic predictors of the use of the internet in these areas. Despite the similar use of the internet by both parents, mothers were found to use the internet more than fathers to seek information in all the family health domains examined. Mothers also reported much higher use of the internet for information relevant to childhood obesity prevention compared with consulting health professionals for advice in this area. Sociodemographic predictors varied between mothers and fathers and the specific topics examined, highlighting the nuances in health information seeking behaviors of parents.

A key finding of interest in this study was that although nearly half of the mothers reported using the internet for information on child feeding and playing, less than 1-fifth reported consulting a health professional on these matters. This is in line

with previous research that suggests mothers of young children prefer digital media because it provides them with free information when they need it the most and at the times when they have opportunities to access it [32]. In contrast, a lack of time and childcare, inconvenient scheduling, and lack of awareness of services or programs have previously been reported as the top barriers to accessing more traditional face-to-face parenting services and programs [18]. This highlights the importance of Web-based digital media as a potential avenue for the delivery of programs to mothers to promote healthy lifestyle behaviors for the whole family. The findings also underscore the importance of understanding the quality of parenting information accessed by parents on the internet.

Despite the almost universal use of the internet by parents in this study, we found, in line with previous research, a much higher proportion of mothers compared with fathers reported using the internet for family health information seeking [22,27]. These differences may reflect the division of responsibility in child rearing, with women more likely to be responsible for cooking and child feeding and gatekeepers for family health and nutrition [33]. However, our own qualitative research [34] shows that fathers believe that they have shared responsibility with respect to the dietary and physical activity behaviors of their young children but felt frustrated by the lack of useable or, at times, conflicting information. This suggests that difficulties accessing or identifying quality Web-based resources may be a barrier for fathers, particularly given the plethora of information available, with limited time exacerbating this issue. Parenting websites also typically target mothers as the primary audience and, therefore, may cater to their interest at the expense of engaging fathers who have been found to have negative attitudes to Web-based parenting resources [27]. Further research is required to explore the best mode of engaging fathers in Web-based parenting resources targeting child lifestyle behaviors.

Our findings regarding the existence of a digital divide in the use of the internet for healthy lifestyle information differed between mothers and fathers. For mothers, maternal age and education were not significant predictors of the use of the internet for information on child feeding and playing. The only significant predictor was employment status, with nonworking mothers nearly twice as likely to search for information on the internet about feeding their child. This may reflect time available outside of paid employment and the more frequent provision of meals to young children among nonworking mothers. In contrast, our findings do suggest a digital divide exists for fathers, with younger, more health conscious, and university-educated fathers more likely to use the internet for information on child health. Again, further research is required to better understand how to engage fathers in digital information relevant to child health and family lifestyle behaviors.

A novel finding in this study was the influence of ethnicity on Web-based information seeking relevant to family health. There was a trend (nonsignificant) for parents born outside Australia to be more likely than Australian-born parents to seek information on the internet related to child feeding (mothers) or child health (fathers). This is in line with 2 qualitative studies that showed that barriers related to language, cultural beliefs, and unfamiliarity with health services drive immigrants to seek postnatal health information primarily from their personal network and the internet [35,36]. Further research is required to understand how ethnic minority groups seek information relevant to family lifestyle behaviors and to better understand the potential for Web-based resources in reaching these parents.

It is important to point out that the multiple regression models only explained a relatively small proportion of the variance in the use of the internet by parents, ranging from as little as 8% of the variance in fathers' use of the internet for their own health to explaining up to 50% of variance in mothers' use of the internet for child feeding information. As is the case with most health behavior research, this suggests that there are many more unmeasured factors influencing the Web-based information seeking behaviors of parents in this study. In support of this, literature in the digital parenting space suggests patterns of health information seeking are complex and reflect differing perceptions of information availability and usefulness, access to health services [37], and comfort with technology [28]. As such, further research is required to unpack the Web-based health information seeking behaviors of parents to maximize the reach and potential use of Web-based parenting resources, particularly in the area of family healthy lifestyle behaviors.

While both mothers and fathers in our study reported being frequent users of Facebook, less than a third of fathers in this study could imagine using Facebook to connect with other parents, compared to nearly 3-quarters of mothers. This may reflect mothers' preferences to connect with other mothers more generally rather than specific preferences for the use of Facebook for this purpose. However, research [15] does suggest that mothers are more likely than fathers to use social media for parenting information as well as social and emotional support for parenting issues. Social media has been proposed as an important emerging context for social influence on mothers' child feeding practices and, thus, a potential avenue for delivery of child obesity prevention interventions [38]. Further research is required to better understand how social media influences obesity-related behaviors in families and how effective interventions could be developed using these platforms.

Strengths and Limitations

This study has a number of strengths and limitations. A key strength of this study was the relatively large sample of parents, enabling predictors of internet use for family health information seeking to be examined separately for mothers and fathers, unique within the existing literature. However, the parents were recruited to participate in a child obesity prevention intervention and, hence, may be more interested in family health behaviors, limiting generalizability. The questions on internet use were not derived from a validated scale or pretested prior to use. As data were collected in 2014-2015, the findings may not reflect parents' current digital health information seeking practices; in particular, we did not examine the use of other social media beyond Facebook. The findings do, however, provide unique insights and ideas for future research.

Conclusions

Our findings support the use of the internet and Facebook as important potential avenues for reaching mothers with information relevant to family lifestyle behaviors, given their frequent use and apparent preference for these sources over consulting health professionals. However, fewer fathers reported using the internet for information relevant to family lifestyle behaviors. Further research is required to understand how best to engage fathers in Web-based parenting resources targeting child obesity prevention, particularly the less educated and less health conscious fathers who are the least likely to seek this information but are likely in greater need of support. Our findings also warrant further investigation of the internet as a potential avenue for reaching ethnic minority parents. The findings from this study, however, only explained a proportion of the variance in the health information seeking behaviors of parents, highlighting the importance of further research in this area to uncover the potential of digital parenting resources and other modes of delivery in supporting healthy lifestyle behaviors in families.

Acknowledgments

The InFANT Extend Program was funded by a World Cancer Research Fund grant (no. 2010/244). KLD is supported by a National Health and Medical Research Council Centre of Research Excellence (APP1057608) grant. KDH is supported by an Australian Research Council Future Fellowship (FT130100637) and an Honorary National Heart Foundation of Australia Future Leader Fellowship (100370), and KK is supported by an Australian Government Research Training Program Scholarship (5488519851).

Conflicts of Interest

None declared.

References

1. UNICEF/WHO/World Bank Group Joint Child Malnutrition Estimates. World Health Organization. 2015. Levels and trends in child malnutrition: Key findings of the 2015 edition URL: http://www.who.int/nutgrowthdb/jme_brochoure2017.pdf [accessed 2018-11-20] [WebCite Cache ID 743lzWkhw]
2. Australian Bureau of Statistics. 2012 Oct 29. 4338.0-Profiles of Health, Australia, 2011-13 URL: http://www.abs.gov.au/ausstats/abs@.nsf/lookup/4338.0main+features12011-13 [WebCite Cache ID 743m6uGL6]
3. Pelone F, Specchia M, Veneziano M, Capizzi S, Bucci S, Mancuso A, et al. Economic impact of childhood obesity on health systems: a systematic review. Obes Rev 2012 May;13(5):431-440. [doi: 10.1111/j.1467-789X.2011.00968.x] [Medline: 22212590]
4. Weng SF, Redsell SA, Swift JA, Yang M, Glazebrook CP. Systematic review and meta-analyses of risk factors for childhood overweight identifiable during infancy. Arch Dis Child 2012 Dec;97(12):1019-1026 [FREE Full text] [doi: 10.1136/archdischild-2012-302263] [Medline: 23109090]
5. Monasta L, Batty G, Cattaneo A, Lutje V, Ronfani L, Van Lenthe FJ, et al. Early-life determinants of overweight and obesity: a review of systematic reviews. Obes Rev 2010 Oct;11(10):695-708. [doi: 10.1111/j.1467-789X.2010.00735.x] [Medline: 20331509]
6. Gibbs BG, Forste R. Socioeconomic status, infant feeding practices and early childhood obesity. Pediatr Obes 2014 Apr;9(2):135-146. [doi: 10.1111/j.2047-6310.2013.00155.x] [Medline: 23554385]
7. Peters J, Sinn N, Campbell K, Lynch J. Parental influences on the diets of 2–5-year-old children: systematic review of interventions. Early Child Development and Care 2012 Jul;182(7):837-857. [doi: 10.1080/03004430.2011.586698]
8. Walsh AD, Cameron AJ, Hesketh KD, Crawford D, Campbell KJ. Associations between dietary intakes of first-time fathers and their 20-month-old children are moderated by fathers' BMI, education and age. Br J Nutr 2015 Sep 28;114(6):988-994. [doi: 10.1017/S0007114515002755] [Medline: 26281910]
9. Walsh AD, Cameron AJ, Crawford D, Hesketh KD, Campbell KJ. Dietary associations of fathers and their children between the ages of 20 months and 5 years. Public Health Nutr 2016 Dec;19(11):2033-2039. [doi: 10.1017/S136898001600077X] [Medline: 27087595]
10. Wake M, Nicholson J, Hardy P, Smith K. Preschooler obesity and parenting styles of mothers and fathers: Australian national population study. Pediatrics 2007 Dec;120(6):e1520-e1527. [doi: 10.1542/peds.2006-3707] [Medline: 18055667]
11. Khandpur N, Blaine R, Fisher J, Davison KK. Fathers' child feeding practices: a review of the evidence. Appetite 2014 Jul;78:110-121. [doi: 10.1016/j.appet.2014.03.015] [Medline: 24667152]
12. Goldfeld SR, Wright M, Oberklaid F. Parents, infants and health care: utilization of health services in the first 12 months of life. J Paediatr Child Health 2003;39(4):249-253. [Medline: 12755928]
13. Dworkin J, Connell J, Doty J. A literature review of parents' online behavior. Cyberpsychology 2013;7(2):A. [doi: 10.5817/CP2013-2-2]
14. Rost J, Johnsmeyer B. Think with Google.: Think Insights with Google Series; 2014. Diapers to Diplomas: What's on the Minds of New Parents URL: https://www.thinkwithgoogle.com/consumer-insights/new-parents/ [accessed 2018-11-20] [WebCite Cache ID 743p0KkiE]

15. Duggan M, Lenhart A, Lampe C, Ellison N. Pew Research Centre. 2015 Jun 16. Parents, social support and outreach on social media URL: http://www.pewinternet.org/2015/07/16/parents-social-support-and-outreach-on-social-media/ [accessed 2018-11-21] [WebCite Cache ID 744D9cx8y]
16. Rhodes A. Royal Children's Hospital National Child Health Poll. Melbourne: The Royal Children's Hospital, Melbourne; 2016. Sources of child health information: What parents use and trust URL: https://www.rchpoll.org.au/polls/sources-of-child-health-information-what-parents-use-and-trust/ [accessed 2018-11-21] [WebCite Cache ID 744Dduwie]
17. Pollard C, Pulker C, Meng X, Kerr D, Scott JA. Who Uses the Internet as a Source of Nutrition and Dietary Information? An Australian Population Perspective. J Med Internet Res 2015 Aug 26;17(8):e209 [FREE Full text] [doi: 10.2196/jmir.4548] [Medline: 26310192]
18. Devolin M, Phelps D, Duhaney T, Benzies K, Hildebrandt C, Rikhy S, et al. Information and support needs among parents of young children in a region of Canada: a cross-sectional survey. Public Health Nurs 2013 May;30(3):193-201. [doi: 10.1111/phn.12002] [Medline: 23586763]
19. Dexter B, Frank S, Seguin L. How and Where Parents of Infants and Young Children Want to Receive Nutrition Information. Can J Diet Pract Res 2016 Dec;77(4):203-205. [doi: 10.3148/cjdpr-2016-020] [Medline: 27763768]
20. Hohman KH, Price SN, Sonneville K, Rifas-Shiman SL, Gortmaker SL, Gillman MW, et al. Can the Internet be used to reach parents for family-based childhood obesity interventions? Clin Pediatr (Phila) 2012 Apr;51(4):314-320 [FREE Full text] [doi: 10.1177/0009922811423310] [Medline: 21997144]
21. Rhodes A. Royal Children Hospital National Child Health Poll. Melbourne: Royal Children's Hospital, Melbourne; 2015. Top ten child health problems: What the public thinks URL: https://www.rchpoll.org.au/polls/top-10-child-health-problems/ [accessed 2018-11-21] [WebCite Cache ID 744E2ByeS]
22. Suarez A, Rodrigo M, Munteon M. Parental activities seeking online parenting support: Is there a digital skill divide? Revista de Cercetare si Interventie Sociala 2016;54(September):36-54.
23. Downing K, Best K, Campbell K, Hesketh KD. Informing Active Play and Screen Time Behaviour Change Interventions for Low Socioeconomic Position Mothers of Young Children: What Do Mothers Want? Biomed Res Int 2016;2016:2139782 [FREE Full text] [doi: 10.1155/2016/2139782] [Medline: 28053979]
24. Vollmer RL, Adamsons K, Gorin A, Foster JS, Mobley AR. Investigating the Relationship of Body Mass Index, Diet Quality, and Physical Activity Level between Fathers and Their Preschool-Aged Children. J Acad Nutr Diet 2015 Jun;115(6):919-926. [doi: 10.1016/j.jand.2014.12.003] [Medline: 25650351]
25. Hall L, Collins CE, Morgan PJ, Burrows TL, Lubans DR, Callister R. Children's intake of fruit and selected energy-dense nutrient-poor foods is associated with fathers' intake. J Am Diet Assoc 2011 Jul;111(7):1039-1044. [doi: 10.1016/j.jada.2011.04.008] [Medline: 21703382]
26. Lima-Pereira P, Bermúdez-Tamayo C, Jasienska G. Use of the Internet as a source of health information amongst participants of antenatal classes. J Clin Nurs 2012 Feb;21(3-4):322-330. [doi: 10.1111/j.1365-2702.2011.03910.x] [Medline: 22093043]
27. He Y, Dworkin J. Cyber dads: What are fathers doing online? Fathering 2015;13(3):219-230. [doi: 10.3149/fth.1303.219]
28. Doty J, Dworkin J, Connell J. Examining digital differences: Parents' online activities. Family Science Review 2012;17(2):18-39.
29. Myers J, Gibbons K, Arnup S, Volders E, Naughton G. Early childhood nutrition, active outdoor play and sources of information for families living in highly socially disadvantaged locations. J Paediatr Child Health 2015 Mar;51(3):287-293. [doi: 10.1111/jpc.12713] [Medline: 25175923]
30. Guerra-Reyes L, Christie V, Prabhakar A, Harris A, Siek KA. Postpartum Health Information Seeking Using Mobile Phones: Experiences of Low-Income Mothers. Matern Child Health J 2016 Dec;20(Suppl 1):13-21 [FREE Full text] [doi: 10.1007/s10995-016-2185-8] [Medline: 27639571]
31. Campbell K, Hesketh K, McNaughton S, Ball K, McCallum Z, Lynch J, et al. The extended Infant Feeding, Activity and Nutrition Trial (InFANT Extend) Program: a cluster-randomized controlled trial of an early intervention to prevent childhood obesity. BMC Public Health 2016 Feb 18;16(1):166 [FREE Full text] [doi: 10.1186/s12889-016-2836-0] [Medline: 26888759]
32. 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] [Medline: 27435182]
33. Tanner C, Petersen A, Fraser S. Food, fat and family: Thinking fathers through mothers' words. Women's Studies International Forum 2014;44(1):209-219.
34. Walsh A, Hesketh K, van der Pligt P, Cameron A, Crawford D, Campbell KJ. Fathers' perspectives on the diets and physical activity behaviours of their young children. PLoS One 2017;12(6):e0179210 [FREE Full text] [doi: 10.1371/journal.pone.0179210] [Medline: 28604810]
35. Nilaweera I, Rowe H, Nguyen H, Burns J, Doran F, Fisher J. Sri Lankan-born women who have given birth in Victoria: a survey of their primary postpartum health-care needs. Aust J Prim Health 2016;22(2):133-139. [doi: 10.1071/PY14067] [Medline: 25703855]
36. Chu CM. Postnatal experience and health needs of Chinese migrant women in Brisbane, Australia. Ethn Health 2005 Feb;10(1):33-56. [doi: 10.1080/1355785052000323029] [Medline: 15841586]
37. Malone M, While A, Roberts J. Parental health information seeking and re-exploration of the 'digital divide'. Prim Health Care Res Dev 2014 Apr;15(2):202-212. [doi: 10.1017/S1463423613000194] [Medline: 23676618]

38. Doub A, Small M, Birch LL. A call for research exploring social media influences on mothers' child feeding practices and childhood obesity risk. Appetite 2016 Apr 01;99:298-305. [doi: 10.1016/j.appet.2016.01.003] [Medline: 26767614]

Abbreviations

InFANT: Infant Feeding, Activity and Nutrition Trial
BMI: body mass index

Edited by G Eysenbach; submitted 25.07.18; peer-reviewed by A While, L Shields; comments to author 05.09.18; revised version received 03.10.18; accepted 04.10.18; published 23.01.19.

Please cite as:
Laws R, Walsh AD, Hesketh KD, Downing KL, Kuswara K, Campbell KJ
Differences Between Mothers and Fathers of Young Children in Their Use of the Internet to Support Healthy Family Lifestyle Behaviors: Cross-Sectional Study
J Med Internet Res 2019;21(1):e11454
URL: http://www.jmir.org/2019/1/e11454/
doi:10.2196/11454
PMID:30674450

Original Paper

How Women Use Digital Technologies for Health: Qualitative Interview and Focus Group Study

Deborah Lupton[1], PhD; Sarah Maslen[2], PhD

[1]News and Media Research Centre, Faculty of Arts & Design, University of Canberra, Bruce, Australia

[2]Faculty of Business, Government and Law, University of Canberra, Bruce, Australia

Corresponding Author:
Deborah Lupton, PhD
News and Media Research Centre
Faculty of Arts & Design
University of Canberra
Building 9
Bruce,
Australia
Phone: 61 04 0307 6791
Fax: 61 02 6201 5381
Email: deborah.lupton@canberra.edu.au

Abstract

Background: A range of digital technologies are available to lay people to find, share, and generate health-related information. Few studies have directed attention specifically to how women are using these technologies from the diverse array available to them. Even fewer have focused on Australian women's use of digital health.

Objective: The Australian Women and Digital Health Project aimed to investigate which types of digital technologies women used regularly for health-related purposes and which they found most helpful and useful. Qualitative methods—semistructured interviews and focus groups—were employed to shed light on the situated complexities of the participants' enactments of digital health technologies. The project adopted a feminist new materialism theoretical perspective, focusing on the affordances, relational connections, and affective forces that came together to open up or close off the agential capacities generated with and through these enactments.

Methods: The project comprised two separate studies including a total of 66 women. In study 1, 36 women living in the city of Canberra took part in face-to-face interviews and focus groups, while study 2 involved telephone interviews with 30 women from other areas of Australia.

Results: The affordances of search engines to locate health information and websites and social media platforms for providing information and peer support were highly used and valued. Affective forces such as the desire for trust, motivation, empowerment, reassurance, control, care, and connection emerged in the participants' accounts. Agential capacities generated with and through digital health technologies included the capacity to seek and generate information and create a better sense of knowledge and expertise about bodies, illness, and health care, including the women's own bodies and health, that of their families and friends, and that of their often anonymous online social networks. The participants referred time and again to appreciating the feelings of agency and control that using digital health technologies afforded them. When the technologies failed to work as expected, these agential capacities were not realized. Women responded with feelings of frustration, disappointment, and annoyance, leading them to become disenchanted with the possibilities of the digital technologies they had tried.

Conclusions: The findings demonstrate the nuanced and complex ways in which the participants were engaging with and contributing to online sources of information and using these sources together with face-to-face encounters with doctors and other health care professionals and friends and family members. They highlight the lay forms of expertise that the women had developed in finding, assessing, and creating health knowledges. The study also emphasized the key role that many women play in providing advice and health care for family members not only as digitally engaged patients but also as digitally engaged carers.

(J Med Internet Res 2019;21(1):e11481) doi:10.2196/11481

KEYWORDS

digital health; websites; online forums; social media; apps; wearable devices; women; Australia; feminist new materialism; qualitative research

Introduction

Over the past three decades, a range of digital technologies for sharing and generating health and medical information has emerged. This began in the mid-1990s following the invention of the internet and the World Wide Web, with the establishment of information websites, email-enabled listservs, blogs, and online discussion forums for lay people to access details about health and medicine and share their own experiences. More recently, mobile devices constantly connected via Wi-Fi, social media platforms, mobile apps, and wearable self-tracking devices have become available to enable people to seek online information at any time and generate and share their own health data and experiences of health care and illness. Recent health care policy emphasizes the importance of active patient engagement with medical expertise, the incorporation of lay expertise into health care delivery, and patient responsibility for understanding and managing their chronic health conditions [1-4]. People are now often expected to be digitally engaged patients [5], actively using digital technologies to seek out information about health and medical issues and manage and promote their health.

The gendered dimensions of the use of digital health technologies have received little attention. Most of these technologies are designed for the universal user: an individual typically assumed to be white, male, and middle class. This can mean that devices do not fit well on female bodies or that menstrual tracking options are not initially included in smartwatch design (as was the case of the first Apple Watch, for example). Female bodily experiences such as menopause or problems with pelvic floor strength tend to be ignored [6-8]. Yet several studies have shown that women are higher users of online health and medical information than men [9,10]. Information websites, online discussion groups, and patient-authored blogs about health and medical topics have helped diverse groups of women, including those with metastatic breast cancer [11] or seeking information about breast cancer [12], women searching for information and support in relation to endometriosis [13], and women with multiple sclerosis [14]. The bulk of research related to women's use of online health sources has focused on digital media for pregnancy and parenting [15]. This research has demonstrated how much women appreciate being able to easily find both medical expert advice and support from other women, who are often located across the globe, experiencing similar health and medical conditions, and willing to share information with each other.

Beyond the specific domains of pregnancy and early parenting, little research has focused on women's use of apps, wearable devices, and social media for health-related purposes. Some studies involving both women and men have identified gendered differences in how people use health and fitness apps. For example, a French study [16] involving interviews with people aged from 20 to over 50 years who had tried using self-tracking diet and fitness apps noted that women were more likely to use diet apps, while more men used fitness apps. Older women were more resistant to using either type of app and more likely to use them for only a short time. Some survey-based research has shown that women tend to use health apps more than men, including studies in Hong Kong [17] and the United States [18]. However, a survey of older Germans (aged over 60 years) found that men were higher users of health apps than women [19], and another recent German survey of adults found little gender difference in health app use [20]. There is very little detailed qualitative research on how women use health and fitness apps. One study has investigated how adolescent girls interested in sports use fitness tracking apps, finding a degree of ambivalence in the participants, particularly in relation to their competitive elements [21].

Only a small number of studies published thus far have addressed Australian women's use of digital health technologies. As with other geographical locations, most Australian-based research on women's use of digital technologies for health-related purposes focuses on pregnancy and parenting. These studies have demonstrated that Australian women experiencing these life stages are keen users of social media and apps to find information, communicate and connect with other mothers, and track their pregnancy and children's development [22-29]. A survey of Australian women who were pregnant or had a child under 3 years found that three-quarters had used a pregnancy app and half had used a parenting app [23]. Little research has investigated other groups of Australian women. One exception is a large survey of Australian young women (aged 18 to 24 years), which found that only 43% had used the internet to search for health information. Those women experiencing stigmatized conditions or symptoms (such as mental health problems) were more likely to have searched online than other participants [30]. Another study of young Australian women diagnosed with a sexually transmissible infection showed that they found both face-to-face and online sources valuable for advice and support [31].

In this paper, we present some key findings from the Australian Women and Digital Health Project. This project is innovative in several ways. First, it included Australian women across a range of ages, education levels, and geographic locations. Second, rather than focusing on specific digital health technologies, it covered the full range currently available to Australian women. Third, it investigated the contextual details of the participants' lived experiences of digital health by using qualitative research methods that invited them to discuss these experiences in detail. Finally, concepts from feminist new materialism were employed to analyze the research materials. This theoretical perspective has not yet been adopted to any great extent in sociocultural analyses of digital health. We employed feminist new materialism to identify the affordances (of both technologies and human bodies), relational connections, and affective forces that came together to open up or close off the agential capacities generated with and through the participants' enactments of digital health.

This approach generated many rich findings, providing novel insights into the women's experiences of different types of digital health technologies. We focus in this paper on the findings that identify how women used health-related information websites, online discussion groups and social media, and apps and wearable devices.

Methods

Research Questions

The Australian Women and Digital Health Project aimed to investigate the following research questions: What digital technologies do women use regularly for health-related purposes, both for themselves and for any others (family members or friends)? Which do they find most and least helpful and useful? What kinds of digital health technologies would they like to see developed in the future? Qualitative methods—semistructured interviews and focus groups—were chosen because they are able to shed light on the situated complexities of the participants' encounters with digital health technologies.

Theoretical Perspective

The theoretical perspective adopted in this project is that of sociomaterialism and particularly feminist new materialism. While sociomaterialist perspectives acknowledge the importance of discourses, imaginaries, social relations, and interactions between people, they direct particular attention to the role played by nonhuman actors in humans' enactments of technologies. This approach recognizes and emphasizes the relational engagements of people with technologies as well as with other people and the dynamic nature of these engagements. Humans and nonhumans (in this case, digital technologies) are viewed as working together to generate agential capacities, a term used in feminist new materialism theory to denote the ways in which people create action and meaning with and through things [32,33]. Empirical analysis is directed at identifying the ways in which "matter comes to matter" [32].

When this theoretical perspective is employed to analyze people's experiences with digital health, it is acknowledged that human bodily sensations, digital technologies, and other humans (for example, family members, members of online communities, and medical practitioners) are involved in complex and ever-changing assemblages [34-36]. The affordances of both technologies and human bodies are brought together. Technological affordances include the uses that are designed into them, while human bodily affordances include enfleshed sensory responses and perceptions, thinking, and memory. The relational connections between actors in these assemblages, both human and nonhuman, and the affective forces generated with and through the intra-actions of these actors [37]—or how people feel emotionally in ways that compel action—contribute to the agential capacities that are created. When adopting a feminist new materialism approach, research materials such as interviews and focus group discussions are analyzed looking for the ways in which people describe their practices referring to these affordances, relational connections, and affective forces and how they work together to open up or close off agential capacities.

Recruitment and Participants

The project comprised 2 separate studies. A total of 66 women participants across the 2 studies were involved in either interviews or focus groups about their use of digital health technologies.

Study 1 involved 3 sets of women living in Canberra, totaling 36 participants. The first set included a total of 11 women who attended an initial community forum advertised among women's community health groups by the Women's Centre for Health Matters, a community-based not-for-profit organization that works in Canberra and surrounding regions to improve women's health. The participants who attended the forum were divided into 2 focus groups, one of which was led by the first author and the other by a staff member from the community center. Participant ages ranged from 28 to 65 years. The forum was used to identify key issues and to test and further develop the questions used for future interviews and focus groups.

Following this forum, another 12 participants (aged 21 to 63 years) were recruited to take part in individual face-to-face interviews. Three further focus groups with a total of 13 women were also conducted. Focus group 1 included 4 women with young children who were part of a support group for mothers living with mental health conditions (aged 25 to 30 years), focus group 2 comprised 6 women (aged 25 to 33 years) with young children, and focus group 3 included 3 women aged in their mid-to-late 50s. Of the total of 36 women involved across these Canberra participant groups, 28 identified their ancestry as Anglo-Celtic and 8 as Asian. Twenty-two participants reported university-level education, while 14 had high school or technical qualifications. These interviews and focus groups were conducted by two research assistants employed on the project. The participants were recruited using the Women's Centre for Health Matter's networks, personal contacts, advertising on relevant Facebook pages (such as those for mothers, people with disabilities, and women's fitness and sporting groups in Canberra), and posters in public places around the city. The meetings took place in a range of locations, including places where the focus group participants usually met, homes, and cafes.

Study 2 involved telephone interviews with 30 women living in various locations around Australia. A market research company was commissioned to recruit the participants and conduct the interviews. Participant information and consent were provided online before the interviews were conducted. This group of participants was recruited using subquotas based on age to ensure a good spread of ages: 10 women aged 18 to 40 years, 10 women aged 41 to 60 years, and 10 women aged 61 years and over. Participants ranged in age from 22 to 74 years. Two-thirds lived in major cities; one-third lived in rural or remote Australia. Twenty participants lived in the state of New South Wales, 4 in Queensland, 5 in Victoria, and 1 in Western Australia. Twenty-five participants described themselves as having Anglo-Celtic ancestry, 1 as western European, 2 as southern European, 2 as Asian, and 1 as middle Eastern. Of this group, 14 reported university qualifications and the remaining 16 participants had high school or technical qualifications.

The same semistructured interview schedule was used with all participants. These questions provided the basis of the interviews and group discussions, but interviewers also probed participants for further comments and explanations of their responses. See Multimedia Appendix 1 for details on the questions asked in the interviews and focus groups.

Ethics Approval and Consent

Ethics approval to conduct this research was granted by the University of Canberra's human ethics research committee (reference number HREC 16-172). All participants were provided with project information and gave their consent to participate. They were all given pseudonyms to protect their anonymity.

Analysis

All the group discussions were audio-taped and transcribed by a professional transcription company. The authors worked together to analyze the transcripts using inductive thematic analysis [38] informed by the feminist new materialism approaches outlined earlier. This involved identifying recurring themes within and across each group discussion by reading and rereading the transcripts, locating the places where the participants talked about the digital information that they accessed from online media, and considering the different ways in which human-technological assemblages came together in the participants' accounts. We focused in particular on identifying the affordances, relational connections, affective forces, and agential capacities that were generated [35]. Our results are organized by topic (overview of information sources, information websites, discussion forums and social media, and health and fitness apps and wearable devices). Verbatim quotations from the discussions were chosen to provide support for the analysis.

Results

Overview of Health Information Sources

The interviews and focus group discussions opened with a contextualizing question asking participants what sources they currently used to access health information. They were specifically asked about every source listed in Table 1, which provides an overview of their responses; they also had the opportunity to list other sources. All of the participants said that they accessed both online sources and face-to-face sources of health information regularly. All referred to visiting doctors and other health care professionals, and the majority noted that in-person interactions with family and friends were also a key source of health information for them. For the most part, traditional media forms such as books were not highly used. However, printed pamphlets did remain influential sources for about half of the participants, particularly as they were available when the women were waiting at doctors' offices for appointments. Other sources of health information were nominated by small numbers of participants (categorized as other in the table). These included a medical phone service, videos, podcasts, information sheet about a medication provided by pharmacist, asking a pharmacist for information, emailed newsletters from groups, magazines, newspapers, television advertisements, and television health programs/documentaries.

The next question asked participants to specify which digital health technologies they currently used. Again, they were specifically asked to respond to each technology listed in Table 2, which gives an overview of their responses.

As Table 2 demonstrates, using a search engine to search for health information online was a universal practice among the participants. Google Search was the only search engine mentioned by the participants: they typically referred to googling or consulting Dr Google when describing this practice. Health and fitness apps were used by over half of the participants. Social media were used less frequently (a third of participants), with Facebook groups most often mentioned as social media sources of information about health. One in five participants was currently using a wearable device for health-related purposes, with Fitbit (Fitbit Inc) fitness trackers and Apple Watches (Apple Inc) the most popular. Small numbers of women said they used digital self-care devices to manage a chronic health condition or exercise games like Wii Fit (Nintendo), while none reported using online physical fitness platforms like Strava. When asked if they used any other digital technology for health-related purposes that had not been listed, 2 said they played mind fitness games online, another 4 women referred to using email to send articles to family members or friends about health issues or being part of email groups set up for health-related topics, and 3 mentioned watching YouTube videos about health issues.

Table 1. Health information sources currently used by participants (N=66).

Source	Value, n (%)
Doctors/other health care providers	66 (100)
Online sources	66 (100)
Friends and family	56 (85)
Pamphlets	31 (48)
Books	12 (18)
Other	13 (20)

Table 2. Digital technologies currently used for health by participants (N=66).

Technology	Value, n (%)
Search engines	66 (100)
Websites	60 (90)
Apps	38 (57)
Social media	50 (33)
Online discussion forums	18 (27)
Wearable device	13 (20)
Exercise games	9 (13)
Self-care devices for chronic illnesses	9 (13)
Physical activity platforms	0 (0)
Other	9 (14)

Health Information Websites

The majority of participants, regardless of their age, geographical location, or level of educational attainment, reported accessing health information websites. It was common for women to say that they went online very regularly: several times a week, in some cases. The definition of health-related information was quite broad in women's accounts. It was interpreted by women to mean baby care; fitness advice; weight loss, diet, healthy eating, or cooking sites; discussion groups or social media groups, as well as sites that offered information about symptoms or medical conditions and treatments. Participants valued the currency of online information compared with traditional printed media, noting that websites were regularly updated.

Many women discussed going online to look for health information for a family member: a child (including adult children), grandchild, their partner, or elderly parent. One explained that she searches for information online on a weekly basis for family members' health conditions.

> *I've got a son who has Asperger syndrome, which is a form of autism, and quite often I'll access information to see if there's anything new coming out. My husband has type 2 diabetes and a heart condition so quite often I'll do research to see if there's anything new coming out about that.* [Susan, 56 years]

Visits to medical practitioners were far less frequent. Susan, for instance, only visits her doctor quarterly on average. The women commented that consulting a doctor was much more time-consuming: it involved taking additional time to make an appointment that could fit into people's schedules and to attend the appointment. They noted that they valued online sources highly because they were accessible at any time, free to use, and could be consulted at length, without time constraints. For some women, the expense of a doctor's consultation was also a factor that deterred them from seeing doctors too often or too readily, if their doctor did not bulk-bill and there were out-of-pocket fees. People with poor access to the type of health services they needed also appreciated online resources. Those women who were caring for young children wanted to avoid the effort required to bring them along to an appointment.

The participants' frequent recourse to websites for health information mostly did not diminish their trust or faith in medical expertise. When they were asked to nominate their major sources of health information, doctors and other health care professionals were mentioned by all the participants. Most women acknowledged that online sources supplemented rather than replaced the expertise and advice of medical practitioners. For example, one commented that she valued the more personal approach she obtains from a doctor's visit. Doctors can examine her closely and draw on their experience in making a diagnosis. This expertise is not offered on the internet.

> *If you're googling stuff, or you're using the internet, or even social media for that matter, they don't know the full story and they can't see what you're talking about—it's just their opinion. But if you go to the doctor, the doctor has studied for this and they know what they're looking for and they know how to deal with it.* [Rachel, 38 years]

The participants also described searching for further information once a diagnosis had been made by a medical professional. Online sources were used to fill gaps in information or explanations of illnesses that participants or their family members received from medical practitioners.

> *For instance, if you go to the doctor and they tell you you've got high blood pressure you would maybe just have a look to get more information about something that the doctor has diagnosed you with. Well with the doctor you're only there for a limited amount of time so they probably can't tell you every single little thing there is to do with it.* [Jodie, 45 years]

Although the participants were highly reliant on websites to find health information, they were not uncritical of the details they found there.

> *Google is most useful technology—I think it's a great place to start, [but] when I do go on to Google I don't go for the first one, I give a bit of time to research all the different sites to try and get a good overview. I*

> *don't believe the first one that might pop up.* [Audrey, 69 years]

Participants were particularly concerned about the accuracy of websites funded by commercial interests. They outlined various strategies they used to determine whether they could trust information they found online. These strategies included looking for government-related websites (such as those run by departments of health), major health charities and organizations, and preferring Australian sources over non-Australian (because they were considered more relevant).

Online Discussion Forums and Social Media

Many women also referred to the value of accessing peer communities such as patient support forums and Facebook groups established to share information about health and medical topics or specific conditions. The key benefits of these sources were the opportunity to share experiences as well as ask advice and find support from other people experiencing similar illnesses or life events. Online discussion forums and social media sites were described by women as providing support and advice from other people in their situation. They particularly valued being able to access a more personalized form of information that provided insights from like-minded others.

> *Just the really interesting things people put on there and real-life experiences and what they've gone through. They give you information, like links to go through, you can either take it or leave it. It's up to you what you get out of it.* [Houda, 45 years]

Several women said that they appreciated the privacy that online forums and social media afforded them, meaning that they could see what other people were saying about health topics and contribute their experiences without needing to reveal their identity.

> *It's more private, yeah, it's definitely more... you know, you are able to look up safe without feeling embarrassed about if you were to talk to somebody about it. It's not judgmental.* [Lara, 33 years]

Online forums and social media groups were sometimes used as a way of finding information based on lived experience that women were having difficulty accessing on medical websites or from their encounters with doctors. One gave an example of finding information for her adult daughter, who has Hashimoto disease (a disorder of the thyroid gland).

> *A lot of googling was discovering what this was and finding out there was a lot of people that had [this condition], and the things that made them sick. It wasn't so much that we were talking on a forum but going in and finding what people had said. Not getting involved in that, but people were able to say, "Okay, I found out that this was blah blah blah and gluten really set me off." And it was very good: very, very good.* [Diane, 56 years]

Other women recounted their experiences of contributing actively to an online forum or social media group. Facebook was the most commonly mentioned social media platform used by the participants for health and medical information and exchanging experiences. Some women, particularly those with young children or who were living with chronic health conditions, said that they were members of more than 10 Facebook groups related to health or parenting. A few women described using Instagram for following fitness influencers and Pinterest for healthy recipes and fitness tips.

For some women, online forums and social media were used in similar ways as health and medical websites: as a first source of information that helped them decide whether they needed to consult a doctor. Those women who used Facebook were sometimes members of numerous special-interest groups devoted to specific health topics, including diseases or medical conditions that they or their children had or nutrition or fitness groups. For example, a focus group participant said that she was caring for a child with allergies and had found an online forum for women in this situation to be a key source of advice and support.

> *For me, I have a forum that's for mums with allergy kids. So I will often just go and read things on there because I find it really useful. I get lots of information that I wouldn't have found otherwise. I do often actively ask questions, but questions that just like an experienced mother would know, not necessarily medical, yeah.*

While the women were mostly very positive about discussion forums and social media they used for health information, advice, and support, some explained that participating in these peer-support communities could take an emotional toll. Some were also conscious of the potential limits of these platforms as an information source and so used them purposefully. A focus group participant who was part of several Facebook groups for mothers of young children described this tension.

> *I think it depends on what you're looking for too, and what you're asking because people have their own opinion and I kind of find they can borderline bully you on Facebook too... There's a fine line between giving your opinion and then like pushing and pushing or being judgmental.*

Health and Fitness Apps and Wearable Devices

For those participants who used health and fitness apps, calorie-tracking apps such as MyFitnessPal and Weight Watchers and physical activity apps for monitoring heart rate, calories burned, and steps taken or apps providing workout or yoga programs and routines were by far the most often mentioned. Other apps nominated included those designed for the following purposes: medication reminders, self-diagnosis, medical insurance, first aid, Medicare (the universal state-funded health care system in Australia), water consumption, child vaccination, booking exercise classes, pelvic floor exercises, and sleep-tracking. Many of these apps were used for information purposes—to look up information about pharmaceuticals, for example, send out reminders, or to generate details about the users' own bodies.

Several women who were using health and fitness self-tracking apps and wearable devices appreciated the better knowledge of their bodies that their practices gave them. They reported struggling with weight loss or attempting to increase their fitness

levels. Being able to use an app or wearable device to closely monitor their bodies helped them exert control over their bodies. It is notable that women in their 50s, 60s, and even 70s reported using apps and wearable devices for these purposes. For example, one uses apps on her iPhone to help her monitor her weight-loss and exercise efforts.

> *I think it just helps you keep track of things a little bit better and easier. Before you had all these apps and things, you never knew how many steps you'd walked or never knew whether your heartbeat was fine or not unless you went to the doctor's or the hospital. So this way it's given you a little bit more control of what your body is telling you. Or knowledge I should say, not so much control, more knowledge about what your body is doing.* [Pearl, 72 years]

The women who used calorie and fitness tracking apps and wearable devices discussed how motivating they were, allowing them to set goals that they could strive to reach. Several women also referred to enjoying the notifications and reminders that apps and wearable devices sent them, again as a way of providing motivation to reaching goals. One woman uses MyFitnessPal to monitor her calorie intake and physical activity.

> *The app lets me know whether I've done well or whether I haven't done well. Because you get a bit complacent with these sorts of things. And it sends me messages as well: "You haven't logged on for a particular amount of time" or "You haven't done enough steps today." It just prompts you with bits of information like that. So it tracks how I'm going and it gives me reminders... Every time those little messages pop up, I think "Oh God, I haven't done that yet!"* [Julie, 51 years]

Among the wearable devices for health and fitness mentioned by participants, Fitbit was the most highly used. Those women who used it described the benefits of being easily able to track their physical activity or their sleep patterns using the automated features of the device.

> *The Fitbit is easy. I don't have to do anything, I'll just wear it and set goals and that's it. It's easy and very measurable. It's quantifiable, there's no point in having goals if you can't quantify it and you can't measure it, well Fitbit makes both of those very easy. I wish I'd had it a lot further back. Even 10 years ago, this was a dream.* [Robyn, 64 years]

Apps to monitor reproductive health and parenting were also popular among the younger women. Several women used period tracking apps such as Clue to help them identify patterns in their menstrual cycle and generate notifications about when their next period could be expected. Women who were pregnant or had young children had used apps for tracking the progress of their pregnancy and finding information (apps such as Ovia and What to Expect When You're Expecting), child vaccination records, infant development monitoring (in particular, the Wonder Weeks app), and parenting advice (for example, Baby Center). These apps served a combination of information provision and generating new data about the infants' health and development. Some of the women in one of the focus groups said that they used apps to track the habits and routines of their infants. They had a conversation in which they talked about apps for tracking feeding, nappy changes, and sleep. They said that they found these apps helpful because "we're still new mums," as one woman put it, and "you're just so tired all the time," another added.

While app and wearable device use could be experienced as empowering and motivating, affective responses such as frustration, irritation, guilt, shame, and disappointment were expressed by some participants who had tried health and fitness apps. Chief among the complaints was poor design: the app did not work properly, kept crashing, or did not sync well to other devices. Some apps demanded too much time from users—this was particularly the case for calorie-counting apps that required users to input data manually each day. Other apps were considered not to be accurate enough. Some women began to find the constant reminders and notifications sent by apps or wearable devices to be overly intrusive. The women in the 2 focus groups comprised of new mothers, for example, noted that while they found self-tracking fitness apps and wearable devices to be helpful before having their babies, they no longer had the time or desire to use them. They commented on the guilt and shame that these technologies could incite in them and also the lack of acknowledgement by the devices or apps that their lives had changed so enormously. One participant observed this of her Garmin smartwatch.

> *I wish that there was a thing that during pregnancy where that I could put in and say "I'm pregnant," because I got those notes that your sleep is really irregular, and I was like, "Because I'm pregnant!" ...It's almost like it's shaming you.*

Discussion

Principal Findings

The findings from the Australian Women and Digital Health Project revealed new insights into how Australian women across a range of ages, geographical locations, and education levels are using the spectrum of digital technologies available to manage and support their health and well-being. There are no previous qualitative studies with which these findings can be directly compared, but they demonstrate that it is not only Australian women in the life stage of pregnancy or early parenting [22-29] who regularly seek information online and actively use social media and apps.

Older digital tools and sources such as search engines, websites, and online discussion forums are often neglected in contemporary discussions of the potential of digital health, while newer digital media such as apps, wearable devices, and social media receive high levels of attention. An important finding from our project is that the older digital media remain very highly used and valued among the participants. Regardless of their sociodemographic characteristics, our participants were avid users of online tools and resources such as search engines and websites to find health and medical information. They referred to valuing the affordances of instant and up-to-date information, the opportunity to search for information privately or anonymously, and the peer support that they could find

online. Those women who used apps and wearable devices appreciated the opportunity to automatically monitor their bodies, engender motivation, and work toward health and fitness goals. The participants were actively working with digital technologies to source, assess, and apply health information and advice to their circumstances. They embraced the ideal of responsible citizenship promulgated by the digitally engaged patient discourse [5].

The participants' accounts revealed that a range of agential capacities was generated with and through women's engagement with digital health technologies. These included the capacity to seek and generate information and create a better sense of knowledge and expertise about bodies, illness, and health care, including the women's own bodies and health, that of their families and friends, and that of their often anonymous online social networks. Affective forces such as motivation, reassurance, control, care, and connection emerged in the participants' accounts. Women described a sense of empowerment from being able to readily access health information online and decide whether or not their concerns about their bodies or the health of family members were warranted, could be dealt with using lay remedies, or required a medical appointment. The participants referred time and again to the capacities of agency and control that using digital health technologies afforded them, including feeling as if they were able to better manage their own health, and in many cases, that of their family members. When the technologies failed to work as expected, these agential capacities were not realized. Women responded with feelings of frustration, disappointment, and annoyance, leading them to become disenchanted with the possibilities of the digital technologies they had tried.

Education levels or geographical location did not appear to play an important role in women's use of digital health. However, age, life stage, and whether a participant was living with a chronic health condition or caring for a child or other family member with such a condition were influential. Reflecting general trends among Australians [39], the use of social media groups, apps, and wearable devices was more common among young and mid-life–aged women compared with those aged 65 years and over. Women with a chronic health condition or caring for a family member with such a condition, as well as those experiencing pregnancy or caring for young children, were among the most avid users of health websites, social media groups, and online forums, seeking peer support and alternative information sources to those offered by doctors as well as orthodox medical advice. However, these women could often be frustrated by the design of apps and wearable devices that did not recognize or cater for their needs in their current life stage or state of health. The capacities of these human-technology assemblages were closed down, as the technological affordances did not align well with their bodily affordances.

While a comparison with Australian men's use of digital health cannot be made given that no previous research has focused on this group, it was notable that many participants reported frequently searching for information on behalf of their male partners as well as other family members. No women mentioned that any family member reciprocated this information sourcing on their behalf. Women have traditionally adopted this gendered role, taking responsibility for protecting and promoting the health of partners, children, and elderly parents [40,41]. These findings show that they are now using digital technologies to perform this type of reproductive citizenship [41] and family caring role.

Trust was an important affective force emerging in women's accounts of how they evaluated online sources of information, what they did with this information, and how they interacted with their medical practitioners. The participants positioned their digital health activities as supplementing rather than replacing expert medical advice. As this suggests, online resources have increased the capacity of lay people to access both medical and lay expertise. Lay expertise was valued for its personalized insights into the everyday worlds of living with a condition and information about which treatments can work best, while medical expertise was valued for its authority, clinical experience, and facilitating access to other resources such as medical testing, drug prescriptions, and specialists. Medical experts tended to be positioned as being able to confirm and validate a self-diagnosis or self-sourced therapy, or alternatively, to allay fears that symptoms were serious. Rather than online health resources competing with expert health professionals, therefore, they were used in a complementary manner by the participants, often in ways that reduced their recourse to face-to-face medical services.

Limitations

A limitation of the project is that while it included quite a diverse range of Australian women, there was an overrepresentation of women with a university education, those living in a metropolitan area, and those from Anglo-Celtic and English-speaking backgrounds. Further research should direct more attention to women who are members of more sociodemographically disadvantaged groups. Their experiences of digital health may be different, and their access to both digital technologies and health services tends to be more limited than women who are more advantaged [42]. In-depth studies focusing specifically on Australian men's uses of digital health, another neglected area of research, would also draw further attention to the gendered nature of the enactment of digital health technologies.

Conclusions

This research emphasized the situated dimensions of the range of digital technologies women used to support and promote their health and that of their family members as well as the nondigital elements of their health experiences and practices. The findings demonstrate that affordances, relational connections, and affective forces were important elements of the agential capacities that were opened up or closed off by women's enactments of digital health. The research drew attention to the nuanced and complex ways in which the participants were engaging with and contributing to online sources of information and using these sources together with face-to-face encounters with doctors and other health care professionals and friends and family members. The findings highlight the lay forms of expertise that women have developed in finding, assessing, and creating health knowledges. The study also emphasizes the key role that many women play in providing

advice and health care for family members. As this study demonstrated, many were actively using digital media to find information not just for themselves but for others. They were performing the roles of both digitally engaged patients [5] and digitally engaged carers.

Acknowledgments

We thank the Women's Centre for Health Matters for their assistance in developing study 1 of the Australian Women and Digital Health Project and helping with recruitment of participants and the community forum. We acknowledge the contribution of Miranda Bruce and Clare Southerton, who worked as participant recruiters and interviewers on study 1, and McCrindle Research, which recruited participants from its panels and conducted and transcribed the study 2 interviews.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview/focus group schedule.

[PDF File (Adobe PDF File), 25KB - jmir_v21i1e11481_app1.pdf]

References

1. Taylor D, Bury M. Chronic illness, expert patients and care transition. Sociol Health Illn 2007 Jan;29(1):27-45 [FREE Full text] [doi: 10.1111/j.1467-9566.2007.00516.x] [Medline: 17286704]
2. Greenhalgh T, Wherton J, Sugarhood P, Hinder S, Procter R, Stones R. What matters to older people with assisted living needs? A phenomenological analysis of the use and non-use of telehealth and telecare. Soc Sci Med 2013 Sep;93:86-94 [FREE Full text] [doi: 10.1016/j.socscimed.2013.05.036] [Medline: 23906125]
3. Hinder S, Greenhalgh T. "This does my head in." Ethnographic study of self-management by people with diabetes. BMC Health Serv Res 2012 Mar 29;12:83 [FREE Full text] [doi: 10.1186/1472-6963-12-83] [Medline: 22458767]
4. Rogers A, Bury M, Kennedy A. Rationality, rhetoric, and religiosity in health care: the case of England's Expert Patients Programme. Int J Health Serv 2009;39(4):725-747. [doi: 10.2190/HS.39.4.h] [Medline: 19927412]
5. Lupton D. The digitally engaged patient: self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Jun 19;11(3):256-270. [doi: 10.1057/sth.2013.10]
6. Epstein D, Lee N, Kang J, Agapie E, Schroeder J, Pina L, et al. Examining menstrual tracking to inform the design of personal informatics tools. 2017 Presented at: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 5-11, 2017; Denver.
7. Almeida T, Comber R, Balaam M. Intimate care as an agenda for change in women's health. 2016 Presented at: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose.
8. Lupton D. The Quantified Self: A Sociology of Self-Tracking. Cambridge: Polity Press; 2016.
9. Bidmon S, Terlutter R. Gender differences in searching for health information on the internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
10. Montagni I, Cariou T, Feuillet T, Langlois E, Tzourio C. Exploring digital health use and opinions of university students: field survey study. JMIR Mhealth Uhealth 2018 Mar 15;6(3):e65 [FREE Full text] [doi: 10.2196/mhealth.9131] [Medline: 29549071]
11. Vilhauer RP. Perceived benefits of online support groups for women with metastatic breast cancer. Women Health 2009 Oct 19;49(5):381-404. [doi: 10.1080/03630240903238719] [Medline: 19851944]
12. Unruh H, Bowen D, Meischke H, Bush N, Wooldridge J. Women's approaches to the use of new technology for cancer risk information. Women Health 2004;40(1):59-78. [doi: 10.1300/J013v40n01_04] [Medline: 15778132]
13. Shoebotham A, Coulson S. Therapeutic affordances of online support group use in women with endometriosis. J Med Internet Res 2016 May 09;18(5):e109 [FREE Full text] [doi: 10.2196/jmir.5548] [Medline: 27160641]
14. Sosnowy C. Practicing patienthood online: social media, chronic illness, and lay expertise. Societies 2014 Jun 13;4(2):316-329 [FREE Full text] [doi: 10.3390/soc4020316]
15. Lupton D, Pedersen S, Thomas G. Parenting and digital media: from the early web to contemporary digital society. Sociology Compass 2016 Aug 02;10(8):730-743. [doi: 10.1111/soc4.12398]
16. Régnier F, Chauvel L. Digital inequalities in the use of self-tracking diet and fitness apps: interview study on the influence of social, economic, and cultural factors. JMIR Mhealth Uhealth 2018 Apr 20;6(4):e101 [FREE Full text] [doi: 10.2196/mhealth.9189] [Medline: 29678807]

17. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. JMIR Mhealth Uhealth 2018 Dec 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]
18. Carroll K, Moorhead A, Bond R, LeBlanc G, Petrella J, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Dec 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
19. 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]
20. Ernsting C, Dombrowski U, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Dec 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
21. Depper A, Howe PD. Are we fit yet? English adolescent girls' experiences of health and fitness apps. Health Sociol Rev 2016 Jul 04;26(1):98-112. [doi: 10.1080/14461242.2016.1196599]
22. Chalklen C, Anderson H. Mothering on Facebook: exploring the privacy/openness paradox. Soc Media Soc 2017 May 17;3(2):1 [FREE Full text] [doi: 10.1177/2056305117707187]
23. Lupton D, Pedersen S. An Australian survey of women's use of pregnancy and parenting apps. Women Birth 2016 Aug;29(4):368-375. [doi: 10.1016/j.wombi.2016.01.008] [Medline: 26874938]
24. Johnson S. Maternal devices, social media and the self-management of pregnancy, mothering and child health. Societies 2014 Jun 13;4(2):330-350 [FREE Full text] [doi: 10.3390/soc4020330]
25. 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]
26. 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] [Medline: 27435182]
27. Lupton D. It just gives me a bit of peace of mind: Australian women's use of digital media for pregnancy and early motherhood. Societies 2017 Sep 15;7(3):25 [FREE Full text] [doi: 10.3390/soc7030025]
28. 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]
29. Hearn L, Miller M, Fletcher A. Online healthy lifestyle support in the perinatal period: what do women want and do they use it? Aust J Prim Health 2013;19(4):313-318. [doi: 10.1071/PY13039] [Medline: 23899373]
30. Rowlands J, Loxton D, Dobson A, Mishra D. Seeking health information online: association with young Australian women's physical, mental, and reproductive Health. J Med Internet Res 2015 May 18;17(5):e120 [FREE Full text] [doi: 10.2196/jmir.4048] [Medline: 25986630]
31. East L, Jackson D, O'Brien L, Peters K. Being diagnosed with a sexually transmitted infection (STI): sources of support for young women. Contemp Nurse 2015;50(1):50-57. [doi: 10.1080/10376178.2015.1013427] [Medline: 26145717]
32. Barad K. Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning. Durham: Duke University Press; 2007.
33. Braidotti R. Posthuman critical theory. In: Banerji D, Paranjape M, editors. Critical Posthumanism and Planetary Futures. Berlin: Springer; 2016:13-32.
34. Fullagar S, Rich E, Francombe-Webb J, Maturo A. Digital ecologies of youth mental health: apps, therapeutic publics and pedagogy as affective arrangements. Soc Sci 2017 Nov 06;6(4):135 [FREE Full text] [doi: 10.3390/socsci6040135]
35. Lupton D. I just want it to be done, done, done! Food tracking apps, affects, and agential capacities. MTI 2018 May 23;2(2):29 [FREE Full text] [doi: 10.3390/mti2020029]
36. Lupton D, Smith G. A much better person: the agential capacities of self-tracking practices. In: Ajana B, editor. Ontologies of Self-Tracking Practices. London: Emerald Publishing; Sep 01, 2018:57-73.
37. Barad K. Diffracting diffraction: cutting together-apart. Parallax 2014 Jul 11;20(3):168-187. [doi: 10.1080/13534645.2014.927623]
38. Ezzy D. Qualitative Analysis. London: Routledge; 2013.
39. Yellow Social Media Report 2018: Part One—Consumers. URL: https://www.yellow.com.au/wp-content/uploads/2018/06/Yellow-Social-Media-Report-2018-Consumer.pdf [accessed 2018-07-04] [WebCite Cache ID 70eQzgHXy]
40. Lupton D. 'It's a terrible thing when your children are sick': motherhood and home healthcare work. Health Sociol Rev 2014 Dec 17;22(3):234-242. [doi: 10.5172/hesr.2013.22.3.234]
41. Lupton D. The reproductive citizen: motherhood and health education. In: Fitzpatrick K, Tinning R, editors. Health Education: Critical Perspectives. London: Routledge; 2014:48-60.
42. Baum F, Newman L, Biedrzycki K. Vicious cycles: digital technologies and determinants of health in Australia. Health Promot Int 2014 Jun;29(2):349-360. [doi: 10.1093/heapro/das062] [Medline: 23144236]

Edited by G Eysenbach; submitted 04.07.18; peer-reviewed by N Coulson, F Régnier; comments to author 08.10.18; revised version received 30.10.18; accepted 30.10.18; published 25.01.19.

Please cite as:
Lupton D, Maslen S
How Women Use Digital Technologies for Health: Qualitative Interview and Focus Group Study
J Med Internet Res 2019;21(1):e11481
URL: http://www.jmir.org/2019/1/e11481/
doi:10.2196/11481
PMID:30681963

Original Paper

Internet Access and Hypertension Management Among the Elderly Population: A Nationally Representative Cross-Sectional Survey in China

Yinzi Jin[1], PhD; Mingxia Jing[2], PhD; Luyu Zhang[3], PhD; Suhang Song[3], PhD; Xiaochen Ma[3], PhD

[1]Department of Global Health, School of Public Health, Peking University, Beijing, China

[2]Department of Public Health, Shihezi University School of Medicine, Xinjiang, China

[3]China Center for Health Development Studies, Peking University, Beijing, China

Corresponding Author:
Xiaochen Ma, PhD
China Center for Health Development Studies
Peking University
Xueyuan Rd 38#, Haidian District
Beijing,
China
Phone: 86 10 82805706
Fax: 86 10 82805706
Email: xma@hsc.pku.edu.cn

Abstract

Background: Hypertension is a rapidly growing epidemic in China. Yet, it remains inadequately controlled, especially in rural areas. The internet has shown potential for better health management in different settings; however, few studies have investigated its role in hypertension management in China.

Objective: This study aims to examine the association between internet access and hypertension awareness, treatment, and control among elderly Chinese adults and to investigate whether the association between internet access and hypertension management differed between those living in urban and rural areas.

Methods: We obtained data from the nationally representative survey of the China Health and Retirement Longitudinal Study in 2011. Hypertension was defined as (1) average systolic blood pressure of ≥140 mm Hg or average diastolic blood pressure of ≥90 mm Hg or (2) currently taking antihypertensive medications. The outcome assessed included hypertension awareness, treatment, and control. The key independent variable was defined as whether one had internet access at home. We performed multivariate logistic regressions for each of the 3 outcomes.

Results: Among 5135 hypertensive respondents (age 62.4 [SD 9.9] years; 2351/5135, 45.78% men), 12.89% (662/5135) had internet access at home. Compared with those who had no internet access, internet access was positively associated with hypertension awareness (odds ratio [OR] 1.36, 95% CI 1.07-1.73) and treatment (OR 1.38, 95% CI 1.09-1.75), but not with control (OR 1.19, 95% CI 0.90-1.58). Internet access reduced urban-rural disparity in hypertension awareness by 9.6% (*P*=.02), treatment by 8.3% (*P*=.05), but not in control. In addition, the moderating effect of internet access on urban-rural disparities in hypertension management was larger among females. The decreased urban-rural disparities were primarily driven by that internet access improved the management level in rural areas.

Conclusions: Despite the low rate of internet access among the elderly population, the internet shows its potential as a platform for achieving better hypertension management in China. Strategies for reducing the disparities in hypertension management and overall disease burden of hypertension among the elderly population might consider the internet as a platform.

(J Med Internet Res 2019;21(1):e11280) doi:10.2196/11280

KEYWORDS

China; health disparity; hypertension; internet

Introduction

Hypertension is a major health problem and largely contributes to morbidity and mortality worldwide [1]. In China, there is a growing epidemic of hypertension among the population as a whole and urban and rural population specifically [2]. Over the past decades, lifestyle and dietary pattern have changed substantially in rural areas owing to China's rapid economic growth and urbanization [3]. As a result, the hypertension prevalence in the rural population is getting close to that of their urban counterparts [3]. However, most health facilities and human resources are located in urban areas [4]. Partially owing to the unequal distribution of health resources, a very pronounced gap exists in hypertension awareness, treatment, and control between urban and rural residence in China [2].

The successful management of hypertension, like optimal control of many other chronic diseases, requires early detection and sufficient compliance to medications [5]. A growing strand of literature has studied the internet as a potential platform for improved disease management. The internet provides patients opportunities for efficient information search and purchase of medication and devices, timely consulting with health providers, as well as real-time data monitoring [6-8]. Positive empirical evidence in developed countries has accumulated in the field of chronic diseases control like diabetes, respiratory disease, and cardiac disease [9-11]. However, 2 key questions remained unanswered; first, does the positive effect on chronic disease management found in other settings apply to hypertension management in China, where access to qualified health workers and effective medications is limited [12]; second, given the large urban-rural disparity in hypertension management in China [13], is the association between internet access and hypertension management differential between urban and rural areas? If so, is internet widening or narrowing existing urban-rural disparities of hypertension management?

To fill this gap, this study uses a nationally representative survey to examine the role of internet access in hypertension management among the elderly population in China. The objectives of this study are 2-fold as follows: (1) to examine the association between internet access and hypertension awareness, treatment and control among elderly Chinese adults and (2) to investigate whether the association between internet access and hypertension management differed between those living in urban and rural areas.

Methods

Data and Sample

Data for this study were obtained from the 2011 round of China Health and Retirement Longitudinal Study (CHARLS 2011). In China, CHARLS is a nationally representative longitudinal survey of adults aged ≥45 years and their spouses, including assessments of social, economic, and health circumstances of community residents [14]. CHARLS 2011 was conducted between June 2011 and March 2012 and included 17,708 participants from 450 communities in 28 provinces. All participants provided written informed consent, and survey protocols were approved by the Peking University Ethics Review Board [15].

This study sample was selected from respondents who completed the survey and provided biomarker information of blood pressure measurement in CHARLS 2011 (n=13,965). The likelihood of nonresponse appeared to be uncorrelated with demographic, socioeconomic status characteristics [15]. Of 13,965 participants, 5260 (37.66%) were found to have hypertension (see the following subsection for detailed measurement of hypertension). Our final analytic sample included 5135 individuals who had hypertension and provided complete data for all study variables (Figure 1).

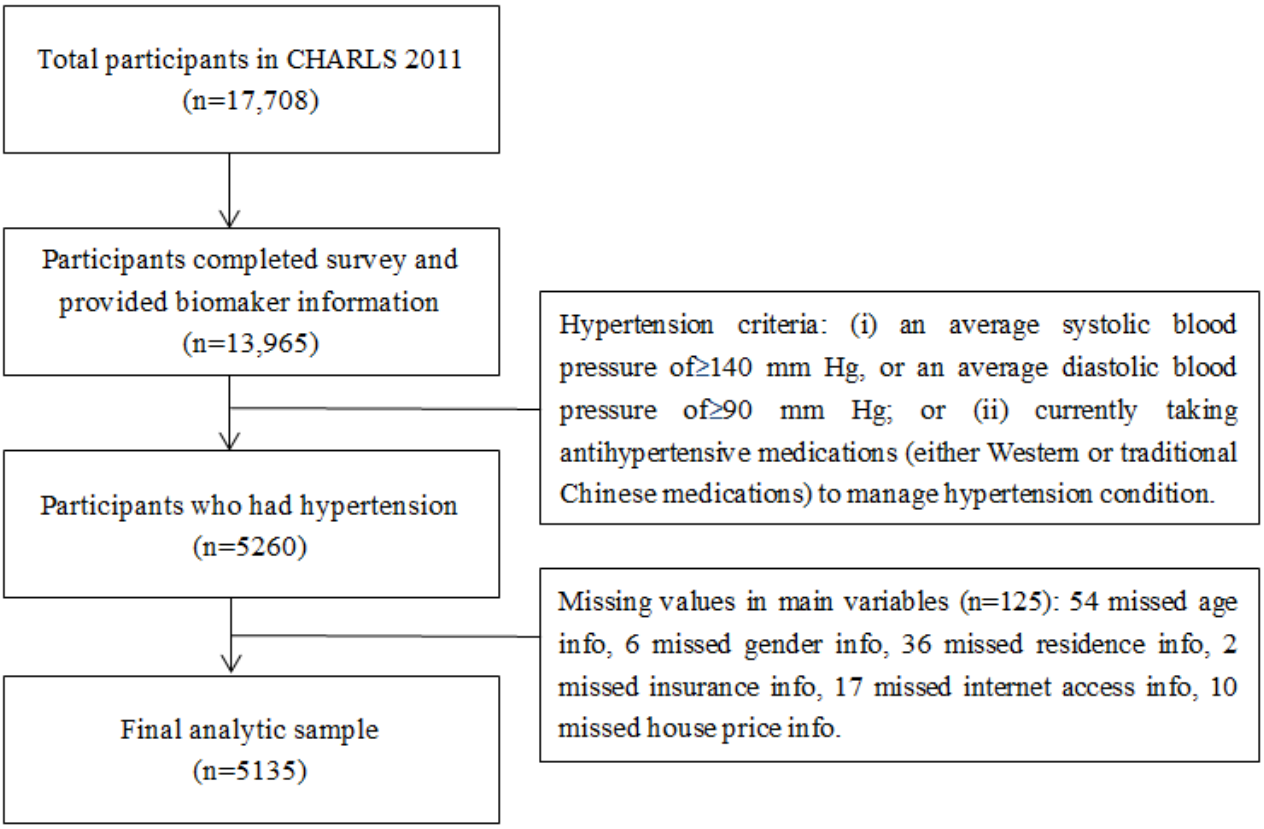
Outcome Measures and Covariates

Blood pressure was measured 3 times at 45-second intervals using an Omron HEM-7200 monitor (Omron, Dalian Co, LTD, Dalian, China) during the daytime. Respondents were asked to relax and remain seated when measured by professionals. The final analytic blood pressure level was calculated by averaging these 3 measures. Hypertension was defined as (1) average systolic blood pressure of ≥140 mm Hg or average diastolic blood pressure of ≥90 mm Hg or (2) currently taking antihypertensive medications (either Western or traditional Chinese medications) to manage hypertension condition.

This study analyzed the following 3 hypertension management variables as outcome variables: (1) *Hypertension Awareness*, defined as an individual with hypertension who reported previous diagnosis of hypertension or simply claimed to have hypertension; (2) *Hypertension Treatment*, defined as an individual with hypertension who claimed to be taking any antihypertensive medications (either Western or traditional Chinese medications); and (3) *Hypertension Control*, defined as an individual with hypertension whose average systolic blood pressure <140 mm Hg and average diastolic blood pressure <90 mm Hg from the blood pressure measures.

Demographic and socioeconomic status covariates for this study included age (45-59, 60-69, and >70), gender (male or female), educational attainment (illiterate, part of primary school, primary school, middle school, high school or above), marital status (married, widowed, separated, divorced, or never married), and insurance status (whether or not uninsured). Family wealth was measured by the market price of the house an individual owned and currently resided. As about 16.38% (841/5135) of our sample did not own a house, their family wealth was coded as zero. Then, a categorical variable of 3 terciles was generated, which cut the sample into lower, middle, and upper tercile based on the value of their family wealth. Lifestyle covariate included smoking behavior (currently smoking or not), comorbidities covariate included the number of co-occurring chronic diseases (whether >3 co-occurring chronic diseases). The key independent variable was defined as whether one had internet access at home. Urban or rural residence was defined as the *Hukou* household registration record of the individual during the survey.

Figure 1. Flowchart of sample selection from the CHARLS 2011.



Statistical Analysis

A descriptive analysis presented the characteristics of study participants among the full sample as a whole and by a subsample of urban and rural areas. Two multivariate logistic regression models were estimated to investigate the role of internet access in hypertension management. Model 1 regressed each outcome variable on internet access and urban residence. Model 2 added an additional regressor of the interaction term between urban residence and internet access. Both models were adjusted for covariates with a $P \leq .05$ in the descriptive analysis. Province fixed effects were added in both models to adjust for unobserved provincial-level factors. The sign of interaction term in Model 2 could be interpreted as whether internet access modified the urban-rural disparity of hypertension management [16,17].

To provide a more intuitive interpretation of the moderating effect of internet access on the urban-rural disparity of hypertension management, we computed the mean and SE of marginal effects for the interaction term in logistic regressions. We repeated our computation using full sample and subsample by gender. In addition, we calculated whether the moderating effect of internet access on urban-rural disparities was primarily driven by the fact that internet changed hypertension management levels in urban or rural areas. All analyses were conducted in Stata 14.1 (StataCorp LP).

Results

Characteristics of Study Participants

Data from 5135 eligible participants with hypertension aged ≥45 years (4000 rural subjects and 1135 urban subjects) were included for analysis. Table 1 provides the characteristics of these participants. Hypertension management levels were low, and among the 5135 participants, awareness, treatment, and control rate were 56.94% (2924), 49.15% (2524), and 20.29% (1042), respectively. Compared with urban participants, rural participants were younger (945/4000, 23.62%, aged ≥70 years in rural areas vs 324/1135, 28.6%, in urban; $P<.001$), comprised more females (2228/4000, 55.70% vs 556/1135, 48.99%; $P<.001$), were less educated (1497/4000, 37.42%, were illiterate in rural areas vs 144/1135, 12.69%, in urban; $P<.001$), were poorer (1160/4000, 29.0%, were in rich tercile of family wealth in rural areas vs 522/1135, 45.99%, in urban; $P<.001$), and had less access to internet at home (329/4000, 8.22% vs 333/1135, 29.34%; $P<.001$), but were less likely to be uninsured (236/4000, 5.90% vs 118/1135, 10.40%; $P<.001$). In terms of hypertension management, rural participants had lower rates of awareness (2165/4000, 54.12%, in rural vs 759/1135, 66.87%, in urban; $P<.001$), treatment (1834/4000, 45.85%, in rural vs 690/1135, 60.79%, in urban; $P<.001$), and control (734/4000, 18.35%, in rural vs 308/1135, 27.14%, in urban; $P<.001$) than their urban counterparts.

Table 1. Characteristics of the study participants.

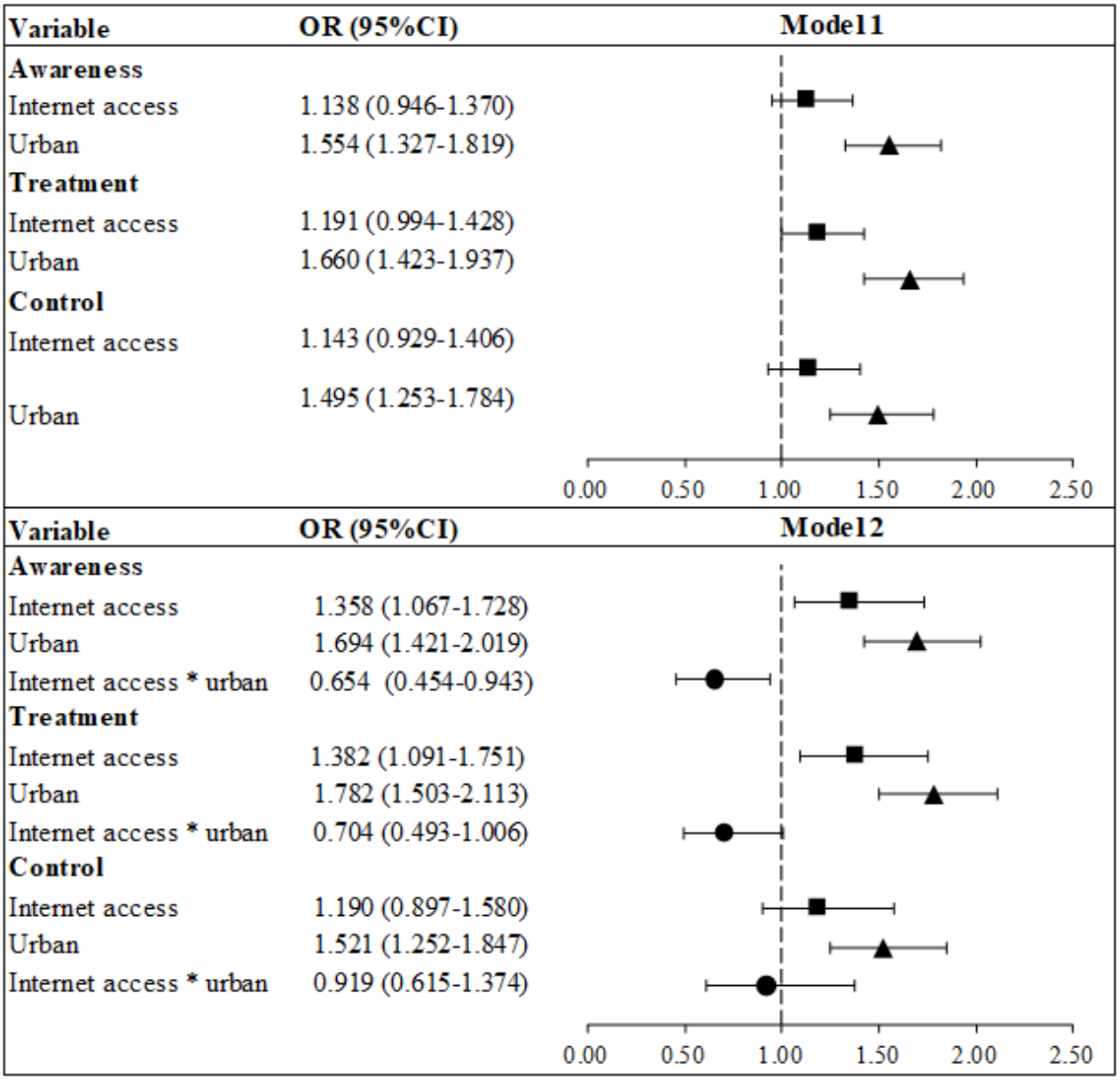
Characteristics	National (n=5135), n (%)	Rural (n=4000), n (%)	Urban, (n=1135), n (%)	*P* value
Covariates				
Age in years				<.001
45-59	2153 (41.93)	1718 (42.95)	435 (38.33)	
60-69	1713 (33.36)	1337 (33.42)	376 (33.13)	
≥70	1269 (24.71)	945 (23.62)	324 (28.55)	
Male sex	2351 (45.78)	1772 (44.30)	579 (51.01)	<.001
Education				<.001
Illiterate	1641 (31.96)	1497 (37.42)	144 (12.69)	
Part of primary school	945 (18.40)	822 (20.55)	123 (10.84)	
Primary school	1139 (22.18)	883 (22.07)	256 (22.56)	
Middle school	913 (17.78)	612 (15.30)	301 (26.52)	
High school or above	497 (9.68)	186 (4.65)	311 (27.40)	
Marital status				.32
Married	4220 (82.18)	3269 (81.72)	951 (83.79)	
Widowed	801 (15.60)	647 (16.18)	154 (13.57)	
Separated, divorced, or never married	114 (2.22)	84 (2.10)	30 (2.64)	
Market price of house				<.001
Poor tercile	1803 (35.11)	1415 (35.38)	388 (34.19)	
Middle tercile	1650 (32.13)	1425 (35.62)	225 (19.82)	
Rich tercile	1682 (32.76)	1160 (29.00)	522 (45.99)	
Uninsured	354 (6.89)	236 (5.90)	118 (10.40)	<.001
>3 co-occurring chronic diseases	1180 (36.61)	1415 (35.38)	465 (40.97)	.001
Currently smoke	2007 (39.08)	1556 (38.90)	451 (39.74)	.61
Internet access at home as key independent variable	662 (12.89)	329 (8.22)	333 (29.34)	<.001
Hypertension management				
Awareness	2924 (56.94)	2165 (54.12)	759 (66.87)	<.001
Treatment	2524 (49.15)	1834 (45.85)	690 (60.79)	<.001
Control	1042 (20.29)	734 (18.35)	308 (27.14)	<.001

Association Between Internet Access and Hypertension Management

Figure 2 plots the odds ratios (ORs) of multivariate logistic regressions. Without adding the interaction term between urban residence and interaction access (Model 1), internet access at home was positively associated with all of the 3 hypertension management outcomes; however, none of the coefficients was statistically significant: hypertension awareness (OR 1.14, 95% CI 0.95-1.37; *P*=.17), treatment (OR 1.19, 95% CI 0.99-1.43; *P*=.06), and control (OR 1.14, 95% CI 0.93-1.41; *P*=.21). In addition, we found significant and notable urban-rural disparities in hypertension awareness (OR 1.55, 95% CI 1.33-1.82; *P*<.001), treatment (OR 1.66, 95% CI 1.43-1.94; *P*<.001), and control (OR 1.50, 95% CI 1.25-1.78; *P*<.001).

After adding the interaction term (Model 2), the positive associations between internet access and hypertension management became statistically significant and with larger coefficient sizes in hypertension awareness (OR 1.36, 95% CI 1.07-1.73; *P*=.01) and treatment (OR 1.38, 95% CI 1.09-1.75; *P*=.007), but not in hypertension control (OR 1.19, 95% CI 0.90-1.58; *P*=.23). In terms of urban-rural disparities, the coefficient sizes were even larger in all 3 outcomes: hypertension awareness (OR 1.69, 95% CI 1.42-2.02; *P*<.001), treatment (OR 1.78, 95% CI 1.50-2.11; *P*<.001), and control (OR 1.52, 95% CI 1.25-1.85; *P*<.001).

Figure 2. Multivariate logistic regressions on multivariate relationship of internet access and urban residence to hypertension management.



Association Between Internet Access and Urban-Rural Disparity of Hypertension Management

The ORs of the interaction term were <1 for each outcome, indicating that internet access reduced the urban-rural gap in hypertension management, although the OR of the interaction term in hypertension control was statistically insignificant. Specifically, the OR of the interaction term between internet access and urban residence in hypertension awareness was 0.65 (95% CI 0.45-0.94; *P*=.02), indicating that the effect of internet access on being aware of hypertensive status for urban residents was 35% smaller than that on rural residents. Similar patterns were also observed for hypertension treatment (OR 0.70, 95% CI 0.49-1.01; *P*=.05) and control (OR 0.92, 95% CI 0.62-1.37; *P*=.68).

Tables 2, 3, and 4 provide a more intuitive interpretation of whether internet access modified the urban-rural disparities on hypertension management by presenting the mean marginal effects of the interaction term between urban residence and internet access. Internet access reduced the urban-rural disparities on hypertension awareness by 9.6% (*P*=.02), hypertension treatment by 8.3% (*P*=.05). We observed no statistically significant change in hypertension control. In addition, these interaction effects were disproportionally and more significantly benefiting female participants; internet access reduced the urban-rural disparities in hypertension awareness by 14.7% (*P*=.01), in hypertension treatment by 14.5% (*P*=.01) among female participants.

The decreased urban-rural disparities in hypertension awareness and treatment associated with internet access were primarily driven by the significant improvement of management level in rural areas among those with internet access at home compared with those without internet access at home. However, the management level remained constant in urban areas between those with and without internet access at home (Tables 5, 6, and 7).

Table 2. Marginal effects of the interaction term between urban residence and internet access for the awareness hypertension.

Population	Coefficient	SE	*P* value
Total (N=5135)	−0.096[a]	0.042	.02
Male (N=2351)	−0.046[a]	0.061	.45
Female (N=2784)	−0.147[a]	0.057	.01

[a]All values adjusted for 6 covariates with a $P \leq .05$ in Table 1: age, gender, educational level, uninsured, the market price of the house, >3 co-occurring chronic diseases. Province fixed effects were also adjusted.

Table 3. Marginal effects of the interaction term between urban residence and internet access for the treatment of hypertension.

Population	Coefficient	SE	*P* value
Total (N=5135)	−0.083[a]	0.042	.05
Male (N=2351)	−0.012[a]	0.062	.83
Female (N=2784)	−0.145[a]	0.058	.01

[a]All values adjusted for 6 covariates with a $P \leq .05$ in Table 1: age, gender, educational level, uninsured, the market price of the house, and >3 co-occurring chronic diseases. Province fixed effects were also adjusted.

Table 4. Marginal effects of the interaction term between urban residence and internet access for the control of hypertension.

Population	Coefficient	SE	*P* value
Total (N=5135)	−0.010[a]	0.036	.79
Male (N=2351)	0.075[a]	0.052	.15
Female (N=2784)	−0.091[a]	0.050	.07

[a]All values adjusted for 6 covariates with a $P \leq .05$ in Table 1: age, gender, educational level, uninsured, the market price of the house, and >3 co-occurring chronic diseases. Province fixed effects were also adjusted.

Table 5. Estimated hypertension awareness by urban residence and internet access.

Variable	Rural[a]	Urban[a]
No Internet	0.43 (0.03)	0.54 (0.02)
Internet	0.49 (0.03)	0.49 (0.03)
P value	.03	.10

[a]All values adjusted for 6 covariates with a $P \leq .05$ in Table 1: age, gender, educational level, uninsured, the market price of the house, and >3 co-occurring chronic diseases. Province fixed effects were also adjusted.

Table 6. Estimated hypertension treatment by urban residence and internet access.

Variable	Rural[a]	Urban[a]
No Internet	0.34 (0.03)	0.47 (0.02)
Internet	0.41 (0.03)	0.44 (0.03)
P value	.02	.27

[a]All values adjusted for 6 covariates with a $P \leq .05$ in Table 1: age, gender, educational level, uninsured, the market price of the house, >3 co-occurring chronic diseases. Province fixed effects were also adjusted.

Table 7. Estimated hypertension control by urban residence and internet access.

Variable	Rural[a]	Urban[a]
No Internet	0.15 (0.03)	0.22 (0.02)
Internet	0.17 (0.02)	0.22 (0.03)
P value	.44	.89

[a]All values adjusted for 6 covariates with a $P \leq .05$ in Table 1: age, gender, educational level, uninsured, the market price of the house, >3 co-occurring chronic diseases. Province fixed effects were also adjusted.

Discussion

Principal Findings

To the best of our knowledge, this study was the first to examine the role of internet access in hypertension management among the elderly population in China. Hypertension is a leading cause of mortality and disability around the world [18]. In China, it accounted for >2.5 million deaths (almost one-third of total deaths) and 15% of total disability-adjusted life-years in 2013, mainly from stroke and ischemic heart disease [18,19]. Over the past decade, the proportion of population connected to the internet has been growing exponentially [20], and researchers in the field of health promotion have been quick to capitalize on the internet to promote changes in health behavior in many settings [21]. Taking advantage of the population-based nationally representative survey CHARLS 2011, we were able to investigate whether the positive effect of the internet on disease management found in other settings still holds in the context of China, where fast economic growth and population aging are happening simultaneously, and more importantly, whether the internet has a moderating effect on the existing urban-rural gap in hypertension management.

In this study, several key findings were highlighted. First, internet access at home was associated with better hypertension management among elderly Chinese adults. Previous studies have found a positive effect of the internet on disease management in different settings [22-26]. For example, one randomized trial proved that internet-based chronic disease self-management program was effective in changing health-related behaviors and improving the health status of patients with chronic diseases [7]. However, few have investigated whether this positive effect still holds in developing context such as China. The positive association between internet access and hypertension management in this study was of practical importance in the sense that China is facing a huge public health crisis of hypertension in which sufficient awareness and compliance to medications are the necessary steps to achieve adequate control.

Second, the finding that internet access was associated with decreased urban-rural disparities in hypertension management was encouraging. The existing gap in hypertension awareness, treatment, and control between urban and rural areas found in this study paralleled that of a recent nationwide population-based epidemiology study of hypertension prevalence and management among Chinese adults [2]. In this study, we were able to use a rich individual-level dataset and examine the association between internet access and urban-rural disparities in hypertension management adjusted for demographic characteristics, socioeconomic status, comorbidity, and lifestyle factors. The magnitudes of 9.6% decrease in hypertension awareness and 8.3% decrease in hypertension treatment were considered to be substantial compared with other studies measuring urban-rural disparities in hypertension management among Chinese patients [13,27]. A well-documented contributor to China's urban-rural disparities in health, including hypertension management, is the unequal distribution of health workforce [28]. Efforts to cope with the shortage of health care workers in rural areas have been slow, especially at primary care level [29]. Our finding provided suggestive evidence that the internet could help to cope with the limited access to quality care in rural areas. More broadly, our results might shed lights to improve health equity in China [30].

Third, a significant and substantial reduction in urban-rural disparity of hypertension awareness and treatment did not translate into the reduction of hypertension control; therefore, improving both is necessary but not sufficient to achieve the optimal control of hypertension [2]. Hypertension control often requires multiple medications, yet in rural China, the availability and affordability to antihypertensive drugs were limited [31]. A recent national study showed that less than one-third of primary health care institutions stocked guideline-recommended and low-cost antihypertensive medications, and unfortunately the deficiencies were even worse in rural areas [31].

Exploring clear mechanisms in explaining the association between internet access and decreased urban-rural disparities in hypertension management was beyond the scope of this study. Nevertheless, we found the association was more notable and statistically significant among female participants. This was in line with a series of studies emphasizing the fact that female patients are more responsive to hypertension management interventions [32,33]. Future studies should focus on understanding channels through which internet access works on improving hypertension management. In addition, decreased urban-rural disparities in hypertension management were primarily driven by the fact that internet access improved management levels in rural areas. This is an encouraging finding, highlighting the importance of improving the hypertension management in rural areas to reduce the burden of hypertension [34]. Furthermore, the results of this study were consistent with previous studies that showed better health management by the internet-based health care models in rural areas and less-developed regions [35,36].

Limitations

This study has several limitations. First, the observational nature of our study limited our ability to draw any causal inference from our findings. The results should not be interpreted as the

effect of internet access on reducing urban-rural disparities in hypertension management. Rather, the association found in this study underscored the need for research to capitalize on new technologies to mitigate the disease burden of hypertension. Second, measuring blood pressure 3 times in the CHARLS study might not lead to the most accurate hypertension diagnosis. The medical literature has confirmed that the 24-hour ambulatory blood pressure monitoring to be the best assessment of hypertension [37]. However, given the large-scale, population-based survey design of CHARLS, the 24-hour ambulatory blood pressure monitoring might not be cost-effective or feasible in implementation. Third, in the 2011 round of CHARLS survey, only 29% and 8% household had internet access in urban and rural areas. This low rate of internet access among our sample population might limit us to generate policy relevance. Although we did not have data of latest internet access among the elderly population, it was reported that internet availability has improved markedly across the country over the past few years [20]. Nonetheless, this lower internet access rate might make our estimates of the association between internet access and reduced urban-rural disparities in hypertension management conservative. Fourth, although the 2.4% nonresponse rate could be considered as low given the large-scale survey of CHARLS, our nonresponse analysis was only limited to comparing hypertension outcome variables between response and nonresponse groups. We have to acknowledge that the nonresponse analysis provided by CHARLS was based on the whole sample, which might not apply to our sample of hypertensive participants. However, because those nonresponses in our study did not complete the survey and only provided biomarker information, we were only to compare the hypertension outcome variables and found nonresponse was not associated with hypertension awareness, treatment, and control (see Multimedia Appendix 1). Last but not least, even if we used a nationally representative survey of the elderly population in China, the generalizability of our results should be limited to the context of China. Given the rapid pace of population aging in China, we believed that our results among elderly adults were of policy importance to that specific population in China.

Conclusions

Hypertension is becoming a public crisis in China and around the globe. Using a nationally representative survey of Chinese elderly, we found that internet access at home was associated with better hypertension management, and internet access reduced the urban-rural gap in hypertension management outcomes. Despite the low rate of internet access among the elderly population, the internet shows its potential as a platform for achieving better hypertension management in China. Longitudinal studies on the internet and hypertension management, because internet facilities have improved significantly in recent years, are needed. Given the growing epidemic of hypertension and the rapid pace of population aging in China, we believe that this study sheds lights on designing polices for achieving optimal hypertension management and health equity in China. Strategies for reducing the disparities in hypertension management and overall disease burden of hypertension might consider the internet as a platform for disease management.

Acknowledgments

The authors would like to acknowledge Jingyuan Xiao (Yale University) for her insightful comments and proofreading of the manuscript. They would also like to acknowledge the CHARLS team for providing data and the training of using the dataset.

This study was supported by Peking University Health Science Center start-up fund (grant no. BMU20160514). The study sponsor has no role in study design, data analysis and interpretation of data, the writing of the manuscript, or the decision to submit the paper for publication.

Authors' Contributions

YJ analyzed the data and revised the manuscript. MJ, LZ, and SS revised the manuscript. XM designed the study, analyzed the data, and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of hypertension management outcomes between response group and non-response group.

[PDF File (Adobe PDF File), 21KB - jmir_v21i1e11280_app1.pdf]

References

1. Olsen MH, Angell SY, Asma S, Boutouyrie P, Burger D, Chirinos JA, et al. A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: the Lancet Commission on hypertension. Lancet 2016 Dec 26;388(10060):2665-2712 [FREE Full text] [doi: 10.1016/S0140-6736(16)31134-5] [Medline: 27671667]

2. Lu J, Lu Y, Wang X, Li X, Linderman GC, Wu C, et al. Prevalence, awareness, treatment, and control of hypertension in China: data from 1·7 million adults in a population-based screening study (China PEACE Million Persons Project). Lancet 2017 Dec 09;390(10112):2549-2558 [FREE Full text] [doi: 10.1016/S0140-6736(17)32478-9] [Medline: 29102084]
3. Li J, Shi L, Li S, Xu L, Qin W, Wang H. Urban-rural disparities in hypertension prevalence, detection, and medication use among Chinese Adults from 1993 to 2011. Int J Equity Health 2017 Dec 14;16(1):50-60 [FREE Full text] [doi: 10.1186/s12939-017-0545-7] [Medline: 28288635]
4. Yip W, Hsiao W, Chen W, Hu S, Ma J, Maynard A. Early appraisal of China's huge and complex health-care reforms. Lancet 2012 Mar 03;379(9818):833-842 [FREE Full text] [doi: 10.1016/S0140-6736(11)61880-1] [Medline: 22386036]
5. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, et al. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA 2003 May 21;289(19):2560-2572 [FREE Full text] [doi: 10.1001/jama.289.19.2560] [Medline: 12748199]
6. Fox SP. Chronic disease and the Internet. Washington, DC: Pew Internet & American Life Project; 2010.
7. Lorig KR, Ritter PL, Laurent DD, Plant K. Internet-based chronic disease self-management: a randomized trial. Med Care 2006 Nov;44(11):964-971 [FREE Full text] [doi: 10.1097/01.mlr.0000233678.80203.c1] [Medline: 17063127]
8. 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 14;15(2):e35 [FREE Full text] [doi: 10.2196/jmir.2439] [Medline: 23410671]
9. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K. Home telehealth for diabetes management: a systematic review and meta-analysis. Diabetes Obes Metab 2009 Oct;11(10):913-930 [FREE Full text] [doi: 10.1111/j.1463-1326.2009.01057.x] [Medline: 19531058]
10. Clark RA, Inglis SC, McAlister FA, Cleland JG, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. BMJ 2007 May 05;334(7600):942-950 [FREE Full text] [doi: 10.1136/bmj.39156.536968.55] [Medline: 17426062]
11. Jaana M, Paré G, Sicotte C. Home telemonitoring for respiratory conditions: a systematic review. Am J Manag Care 2009 May;15(5):313-320 [FREE Full text] [Medline: 19435399]
12. National Health and Family Planning Commission. Health statistics yearbook. Beijing: Ministry of Health; 2012.
13. Hou X. Urban-rural disparity of overweight, hypertension, undiagnosed hypertension, and untreated hypertension in China. Asia Pac J Public Health 2008;20(2):159-169 [FREE Full text] [doi: 10.1177/1010539507312306] [Medline: 19124310]
14. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol 2014 Feb;43(1):61-68 [FREE Full text] [doi: 10.1093/ije/dys203] [Medline: 23243115]
15. Zhao Y, Strauss J, Yang G, Giles J, Hu P, Hu Y, et al. China health and retirement longitudinal study? 46-2012 national baseline users? guide. Beijing: Peking University; 2011.
16. Norton EW, Ai C. The Stata Journal. 2004. Computing interaction effects and standard errors in logit and probit models URL: https://www.researchgate.net/profile/Paul_Louangrath/post/How_can_I_ENTER_a_4-way_interactions_of_variables_in_blocks_of_the_hierarchical_multiple_regression_analysis_in_SPSS/attachment/59d644d2c49f478072ead786/AS:273819644694528@1442295099036/download/INTERACTION-1.pdf [WebCite Cache ID 74GZpZAMt]
17. Buis ML. The Stata Journal. 2010. Stata tip 87: Interpretation of interactions in non-linear models URL: http://ageconsearch.umn.edu/bitstream/152319/2/sjart_st0194.pdf [accessed 2018-11-28] [WebCite Cache ID 74Ga4Uy11]
18. GBD 2013 Risk Factors Collaborators, Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, et al. . [doi: 10.1016/S0140-6736(15)00128-2] [Medline: 26364544]
19. Wang L, Kong L, Wu F, Bai Y, Burton R. Preventing chronic diseases in China. Lancet 2005 Nov 19;366(9499):1821-1824 [FREE Full text] [doi: 10.1016/S0140-6736(05)67344-8] [Medline: 16298221]
20. Kulkarni R, Malouin RA. State of Telehealth. N Engl J Med 2016 Dec 06;375(14):1399-1400 [FREE Full text] [doi: 10.1056/NEJMc1610233] [Medline: 27705258]
21. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
22. Parati G, Torlasco C, Omboni S, Pellegrini D. Smartphone Applications for Hypertension Management: a Potential Game-Changer That Needs More Control. Curr Hypertens Rep 2017 Jun;19(6):48 [FREE Full text] [doi: 10.1007/s11906-017-0743-0] [Medline: 28501982]
23. Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, Web communication, and pharmacist care on hypertension control: a randomized controlled trial. JAMA 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: 10.1001/jama.299.24.2857] [Medline: 18577730]
24. Yatabe MS, Yatabe J, Asayama K, Staessen JA, Mujaj B, Thijs L, et al. The rationale and design of reduction of uncontrolled hypertension by Remote Monitoring and Telemedicine (REMOTE) study. Blood Press 2018 Dec;27(2):99-105 [FREE Full text] [doi: 10.1080/08037051.2017.1406306] [Medline: 29172715]

25. Lu JF, Chen CM, Hsu CY. Effect of home telehealth care on blood pressure control: A public healthcare centre model. J Telemed Telecare 2017 Jan 01:1357633X17734258 [FREE Full text] [doi: 10.1177/1357633X17734258] [Medline: 29041841]
26. AbuDagga A, Resnick HE, Alwan M. Impact of blood pressure telemonitoring on hypertension outcomes: a literature review. Telemed J E Health 2010 Sep;16(7):830-838 [FREE Full text] [doi: 10.1089/tmj.2010.0015] [Medline: 20815751]
27. Wang J, Zhang L, Wang F, Liu L, Wang H, China National Survey of Chronic Kidney Disease Working Group. Prevalence, awareness, treatment, and control of hypertension in China: results from a national survey. Am J Hypertens 2014 Nov;27(11):1355-1361 [FREE Full text] [doi: 10.1093/ajh/hpu053] [Medline: 24698853]
28. Anand S, Fan VY, Zhang J, Zhang L, Ke Y, Dong Z, et al. China's human resources for health: quantity, quality, and distribution. Lancet 2008 Nov 15;372(9651):1774-1781 [FREE Full text] [doi: 10.1016/S0140-6736(08)61363-X] [Medline: 18930528]
29. Yip W, Hsiao W. Harnessing the privatisation of China's fragmented health-care delivery. Lancet 2014 Aug 30;384(9945):805-818 [FREE Full text] [doi: 10.1016/S0140-6736(14)61120-X] [Medline: 25176551]
30. Tang S, Meng Q, Chen L, Bekedam H, Evans T, Whitehead M. Tackling the challenges to health equity in China. Lancet 2008 Oct 25;372(9648):1493-1501 [FREE Full text] [doi: 10.1016/S0140-6736(08)61364-1] [Medline: 18930531]
31. Su M, Zhang Q, Bai X, Wu C, Li Y, Mossialos E, et al. Availability, cost, and prescription patterns of antihypertensive medications in primary health care in China: a nationwide cross-sectional survey. Lancet 2017 Dec 09;390(10112):2559-2568 [FREE Full text] [doi: 10.1016/S0140-6736(17)32476-5] [Medline: 29102087]
32. Lustria ML, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011 Sep;17(3):224-243 [FREE Full text] [doi: 10.1177/1460458211414843] [Medline: 21937464]
33. DiMaggio PH. From the 'Digital Divide' to 'Digital Inequality'tudying Internet Use As Penetration Increases. Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University; 2001:1-23.
34. Hou Z, Meng Q, Zhang Y. Hypertension Prevalence, Awareness, Treatment, and Control Following China's Healthcare Reform. Am J Hypertens 2016 Apr;29(4):428-431 [FREE Full text] [doi: 10.1093/ajh/hpv125] [Medline: 26232034]
35. Wang T, Li J, Zhu C, Hong Z, An D, Yang H, et al. Assessment of Utilization and Cost-Effectiveness of Telemedicine Program in Western Regions of China: A 12-Year Study of 249 Hospitals Across 112 Cities. Telemed J E Health 2016 Dec;22(11):909-920 [FREE Full text] [doi: 10.1089/tmj.2015.0213] [Medline: 27314300]
36. Piette JD, Mendoza-Avelares MO, Milton EC, Lange I, Fajardo R. Access to mobile communication technology and willingness to participate in automated telemedicine calls among chronically ill patients in Honduras. Telemed J E Health 2010 Dec;16(10):1030-1041 [FREE Full text] [doi: 10.1089/tmj.2010.0074] [Medline: 21062234]
37. O'Brien E, Coats A, Owens P, Petrie J, Padfield PL, Littler WA, et al. Use and interpretation of ambulatory blood pressure monitoring: recommendations of the British hypertension society. BMJ 2000 Apr 22;320(7242):1128-1134 [FREE Full text] [Medline: 10775227]

Abbreviations

CHARLS: China Health and Retirement Longitudinal Study
OR: odds ratio

Edited by G Eysenbach; submitted 17.06.18; peer-reviewed by D Nault, Y Yang; comments to author 06.09.18; revised version received 30.10.18; accepted 01.11.18; published 31.01.19.

Please cite as:
Jin Y, Jing M, Zhang L, Song S, Ma X
Internet Access and Hypertension Management Among the Elderly Population: A Nationally Representative Cross-Sectional Survey in China
J Med Internet Res 2019;21(1):e11280
URL: http://www.jmir.org/2019/1/e11280/
doi:10.2196/11280
PMID:30702439

Original Paper

Technical Support by Smart Glasses During a Mass Casualty Incident: A Randomized Controlled Simulation Trial on Technically Assisted Triage and Telemedical App Use in Disaster Medicine

Andreas Follmann[1], MD; Marian Ohligs[1,2], MSc; Nadine Hochhausen[1], MD; Stefan K Beckers[1,3], MD; Rolf Rossaint[1], MD; Michael Czaplik[1,2], MD, PhD

[1]Medical Technology Section, Department of Anaesthesiology, University Hospital RWTH Aachen, Aachen, Germany

[2]Docs in Clouds GmbH, Aachen, Germany

[3]Medical Direction, Emergency Medical Service, City of Aachen, Aachen, Germany

Corresponding Author:
Andreas Follmann, MD
Medical Technology Section
Department of Anaesthesiology
University Hospital RWTH Aachen
Pauwelsstraße 30
Aachen, D-52074
Germany
Phone: 49 2418036219
Email: afollmann@ukaachen.de

Abstract

Background: To treat many patients despite lacking personnel resources, triage is important in disaster medicine. Various triage algorithms help but often are used incorrectly or not at all. One potential problem-solving approach is to support triage with Smart Glasses.

Objective: In this study, augmented reality was used to display a triage algorithm and telemedicine assistance was enabled to compare the duration and quality of triage with a conventional one.

Methods: A specific Android app was designed for use with Smart Glasses, which added information in terms of augmented reality with two different methods—through the display of a triage algorithm in data glasses and a telemedical connection to a senior emergency physician realized by the integrated camera. A scenario was created (ie, randomized simulation study) in which 31 paramedics carried out a triage of 12 patients in 3 groups as follows: without technical support (control group), with a triage algorithm display, and with telemedical contact.

Results: A total of 362 assessments were performed. The accuracy in the control group was only 58%, but the assessments were quicker (on average 16.6 seconds). In contrast, an accuracy of 92% (P=.04) was achieved when using technical support by displaying the triage algorithm. This triaging took an average of 37.0 seconds. The triage group wearing data glasses and being telemedically connected achieved 90% accuracy (P=.01) in 35.0 seconds.

Conclusions: Triage with data glasses required markedly more time. While only a tally was recorded in the control group, Smart Glasses led to digital capture of the triage results, which have many tactical advantages. We expect a high potential in the application of Smart Glasses in disaster scenarios when using telemedicine and augmented reality features to improve the quality of triage.

(J Med Internet Res 2019;21(1):e11939) doi:10.2196/11939

KEYWORDS

augmented reality; disaster medicine; emergency medical service physician; mass casualty incident; Smart Glasses; telemedicine; triage

Introduction

Terrorist attacks, natural disasters, and major traffic accidents present challenges to emergency physicians around the world. The lack of information about those affected and injured and an initial difference between available and necessary resources require a quick overview of the overall situation. In disaster medicine, various strategies exist for assisting the management of mass casualties and prioritizing injured persons according to the need and available resources. Serious injuries require urgent treatment; immediate life-saving measures may be required. Slightly injured persons have to be cared for, but their transport from the damage area can be postponed owing to limited resources; this results in the need of prioritization of treatment and transport to a hospital—the procedure of triage [1]. The triage category is encoded in color and recorded with an individual identification number on a so-called "injured attachment card."

In many countries around the world, different triage algorithms are established to help rescue service personnel and emergency physicians in assigning casualties to one of the triage categories [2]. Algorithms are different for doctors and nonmedical staff who perform the so-called "pretriage"; this is important for many reasons, including providing an overview of the reclaiming of additional personnel. The Primary Ranking for Initial Orientation in Rescue service (PRIOR) algorithm is a pretriage algorithm often used in Germany [3]. Qualitatively, patient consciousness, breathing, and circulation are addressed, which results in the case of nonpathological evaluation in further questions for individual triage; this can be divided into 3 categories as follows: severely injured with immediate treatment priority (category red or I); severely injured with appropriate transport priority (category yellow or II); and easily injured or uninjured (category green or III). However, various triage algorithms are often used incorrectly or not at all [4-6]; this results in incorrect assignments of triage categories and wrong prioritization. Subsequently, scarce resources cannot be used correctly, appropriate treatment priorities are neglected, and treatments delayed [7,8]. All of these are in contradiction with the principles of triage in disaster medicine [9]. From this, it can be concluded that technical support for triage is urgently needed. Therefore, we used the augmented reality and the potential of data glasses for this important task.

To provide technical support for this important phase in the case of mass casualty incidents (MCI), we have developed a triage app running on Smart Glasses (Recon Jet, Recon Instruments, Vancouver, BC, Canada) within the framework of the project Augmented Disaster Medicine (AUDIME), which was financed by the German Federal Government. The AUDIME project aims to develop new high-tech strategies, including apps for Smart Glasses, for the technical support of an MCI [10]. Data are displayed on a small monitor on the Smart Glasses, the video stream of an integrated camera is used for telemedical support. With simple operating gestures provided on an optical touchpad on the Smart Glasses, menu items can be selected, and simple manual entries can be made.

This study aims to evaluate the feasibility of various technical methods for triage support using Smart Glasses. The average duration of a screening process and the accuracy of the assignment to one of the listed triage categories are the primary target parameters.

Methods

Android App

For Smart Glasses, an app was developed together with our project partner (Tech2go GmbH, Hamburg, Germany) for use on Android devices. Using several menu levels, information can be displayed to task forces, and appropriate support can be offered. Triage support was achieved through the display of PRIOR as an example of one of the various triage algorithms (Figure 1). With simple hand movements above the optical touchpad, the decision tree can be processed, and the result of the triage is displayed. The result is recorded digitally and assigned to the individual ID of the patient appendix card through a photo.

Telemedical Support

Another method for investigating technical support provided through Smart Glasses is telemedicine. In Aachen, Germany, this has been used for many years in individual medical emergency care in the routine prehospital rescue service [11-15]. A telemedical-connected emergency medical service (EMS) physician can offer medical assistance to the ambulance staff for making difficult decisions. Telemedicine has not yet been used in disaster medicine. Therefore, the camera of the Smart Glasses was used in this project to provide information about the MCI site through a live video transmission to a tele-EMS-physician. The physician was thus able to gain an impression of the situation and carry out triage collaboratively with the on-site team and assign each patient to a triage category. For this purpose, a separate tele-senior-EMS-physician software (Tele-LNA, Docs in Clouds GmbH, Aachen, Germany) was developed. Again, a digital recording of the triage results was retained. The audio connection was achieved through Voice over Internet Protocol. For better voice quality, a Bluetooth-connected headset was used. Optionally, a battery pack can be connected to optimize the battery capacity of the Smart Glasses even for longer periods of use (Figure 2).

The function of the AUDIME system requires a local Wi-Fi network, which can be spanned by local access points on site; although this technology is still unusual at a deployment site, we assume that many rescue vehicles in the future will have such technology installed to ensure interoperability between rescue and medical technology. Alternative concepts, such as the use of the mobile network, are also conceivable. Using a secure data connection, all collected data are transferred to an information integration layer of the AUDIME server, analyzed, and made available for use on Smart Glasses or other devices, such as tablet personal computers.

Figure 1. Question regarding unconsciousness is asked (“yes” or “no”) on the Smart Glasses screen.

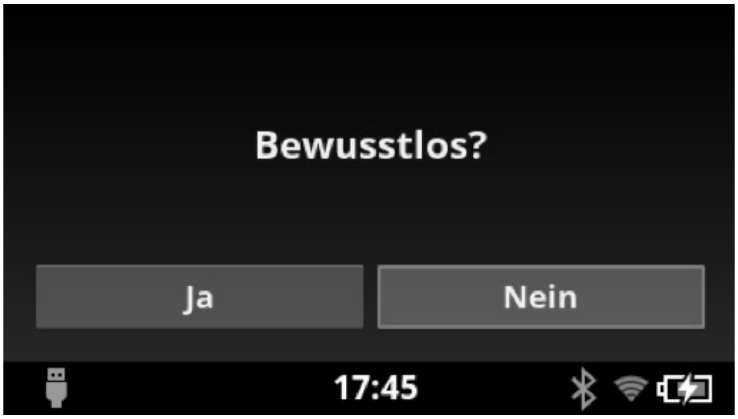
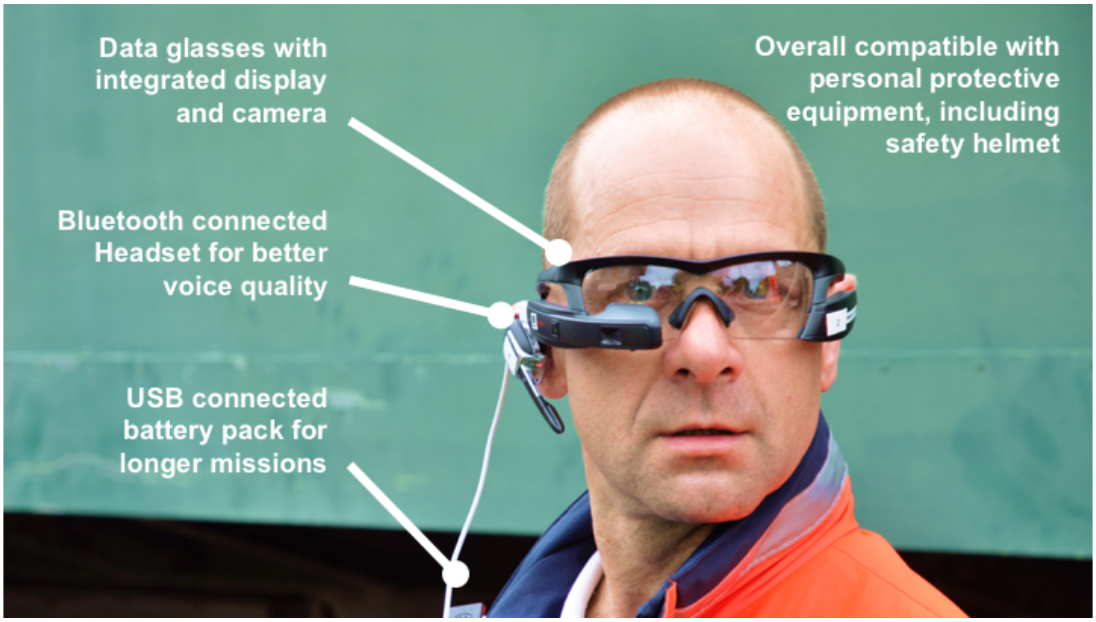


Figure 2. The subcomponents of the triage technical support system.



Study Design

For evaluation in a randomized triage study, a simulation was set up on the grounds of the Fire Brigade and Rescue Training Center in Frankfurt with the approval of the Local Ethics Committee (Aachen, Germany, EK 185/17), which represented an explosion in a row of residential buildings. A total of 12 professional actors mimed patients with different injuries (Table 1). Two of each injury were of identical make-up. The actors were required to exaggerate the case once and then understate it once.

A total of 31 paramedics were available as subjects for the simulation study. Each had sufficient experience of at least 2 years being in practice in EMS. The inclusion criteria for participation, in addition to requisite vocational training, was sufficient eyesight with contact lenses or without visual aids, as the Smart Glasses do not accommodate the wearing of prescription or reading glasses; therefore, this was considered an exclusion criterion. Participants were randomly assigned to 1 of the 3 groups as follows: group 1 was asked to perform an individual triage without further aids and document the results in a tally (control group); group 2 was asked to use the PRIOR algorithm provided by the Smart Glasses app; and group 3 was asked to contact a tele-senior EMS physician to collaboratively carry out the triage of an injured through a video streaming through the integrated camera of the data glasses. Group assignments were unknown to subjects in advance (blinding). The tele-EMS physician had specialist training in anesthesia, intensive care, and emergency medicine, as well as many years of experience as a regular EMS physician.

To ensure the same level of knowledge regarding triage algorithms, all paramedics, including the control group, were trained on the PRIOR algorithm half an hour prior to the simulation (no one knew it before). In addition, all paramedics were provided with the PRIOR algorithm as a pocket card. Subsequently, they were assigned to the method they should use for the triage depending on the study group to which they belonged. The size of each group was based on the estimated time spent on the experiences of the first pilot study.

Table 1. The overview of injury patterns of simulation patients.

Number	Injury	Type of simulation	Triage category
1	Unconsciousness	N/A[a]	I/red
2	Unconsciousness	N/A	I/red
3	Open lower leg fracture	Understated	II/yellow
4	Open lower leg fracture	Exaggerated	II/yellow
5	Piling lower leg	Understated	II/yellow
6	Piling lower leg	Exaggerated	II/yellow
7	Deep thigh wound, no bleeding	Understated	II/yellow
8	Deep thigh wound, no bleeding	Exaggerated	II/yellow
9	Cuts face	Understated	III/green
10	Cuts face	Exaggerated	III/green
11	Head wound	Understated	III/green
12	Head wound	Exaggerated	III/green

[a]Not applicable.

Successively, subjects completed the screening process according to their group assignment. In each case, they were unobtrusively accompanied by an auxiliary who documented the duration of triage and the selected category. Then, the primary target parameters were assessed—duration and accuracy of triage. If the selected triage category did not agree with the correct category, so-called “major deviations” were defined for selected green (III) instead of red (I) or red (I) instead of green (III) category. Subsequently, paramedics of the 2 technical supported groups completed a usability questionnaire. In addition, the feelings of safety of all participating teams and their individual opinions on the executed triage processes are inquired. The data from the subject survey were defined as further outcome parameters.

Statistical Analysis

Data analysis was performed on the nonparametric distribution of primary outcome parameters using the Mann-Whitney U test for independent samples (significance level, P=.05). We used SPSS Statistics software, version 23 (IBM) for statistical evaluation and data backbone. All data are stated as median, interquartile range (IQR), or minima to maxima, respectively. The evaluation of the secondary outcome parameters was descriptive.

Results

User Results

In total, 362 individual triages were performed by 31 paramedics (Table 2); 20 paramedics conventionally triaged 240 patients with manual coverage of the triage category (control group), 7 paramedics triaged a total of 84 patients with an indication of the PRIOR algorithm in the Smart Glasses, and 4 paramedics performed a total of 38 triages along with a tele-senior EMS physician.

Duration of Triage

While the average triage time in the control group (conventional triage with manual coverage of the triage category in a tally) was 16.6 seconds (IQR 11.3-23.6), screening with technical support was longer. The average triage time when using the PRIOR in the Smart Glasses was 37.0 seconds (IQR 28.7-40.4; P=.001), a triage along with the tele-senior EMS physician averaged 35.0 seconds (IQR 31.7-41.1; P=.01). Thus, a triage with technical support and digital capture took markedly longer than conventional triage (Figure 3). The figure includes time required per triage type: Conventional triage, Triage with the PRIOR display in data glasses, and Triage with tele-senior EMS physician by the integrated camera of data glasses. Only technically supported procedures involve a digital capture of the category. Looking at the individual triage categories, the increased time requirements of the technically supported groups, especially in category II and III, were observed (Figure 4).

Quality of Triage

The correct assignment to a triage category was the primary outcome parameter. In the conventional triage, subjects reached an average accuracy of 58% (IQR 33%-75%) of the triage results. The quality of the triages could, however, be markedly increased with technical support. The accuracy was significantly increased to 92% (IQR 75%-92%) for the triage with the PRIOR Smart Glasses app (P=.04) and to 90% (IQR 82%-98%; P=.01) with tele-senior-EMS-physician assistance by data glasses (Figure 5). Deviations of 2 category ranks were found only in the control group. The amount of these “major deviations” was 8% (IQR 0%-15%).

With regard to the outcome parameters, both, the usability and the sense of safety, showed sufficient acceptance among subjects. Of participants who were supported by either the triage algorithm or a telemedical contact, 73% (8/11) stated good or very good usability of the Smart Glasses. In addition, most subjects confirmed compatibility with the personal protective equipment in the event of a disaster. In the questionnaire’s free

text, several subjects added that they felt markedly safer during the triage owing to the technical support. Only one test person regarded the assistance by a tele-senior EMS physician during the triage as superfluous, as “one does not need a doctor for triage.” This subject felt the doctor’s contact was intrusive, and he would prefer to do triage alone in the future.

Table 2. The demographic profile.

Group	Control group (n=20)	Triage with PRIOR[a] service display (n=7)	Triage with tele-senior emergency medical service physician (n=4)
Gender, n (%)			
Females	4 (20)	1(14)	0 (0)
Males	16 (80)	6 (86)	4 (100)
Age (years), mean (range)	32.5 (21-50)	34.1 (28-40)	26.8 (24-31)

[a]PRIOR: Primary Ranking for Initial Orientation in Rescue.

Figure 3. Time (CI) required per triage type. Circles with numbers denote outliers. PRIOR: Primary Ranking for Initial Orientation in Rescue service; EMS: emergency medical service.

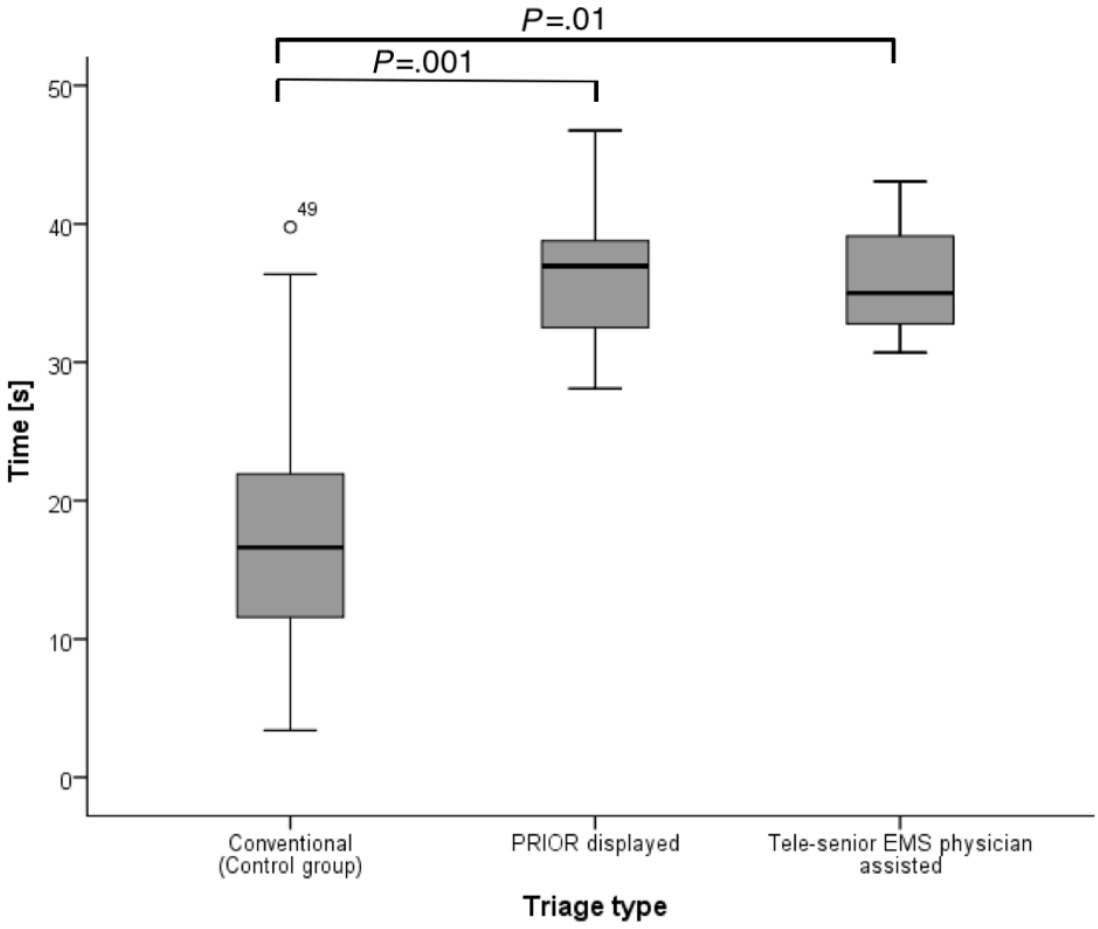


Figure 4. Time (CI) required for a triage in each individual category (from I, seriously injured to III, slightly injured) per each triage group Circles with numbers denote outliers. PRIOR: Primary Ranking for Initial Orientation in Rescue service; EMS: emergency medical service.

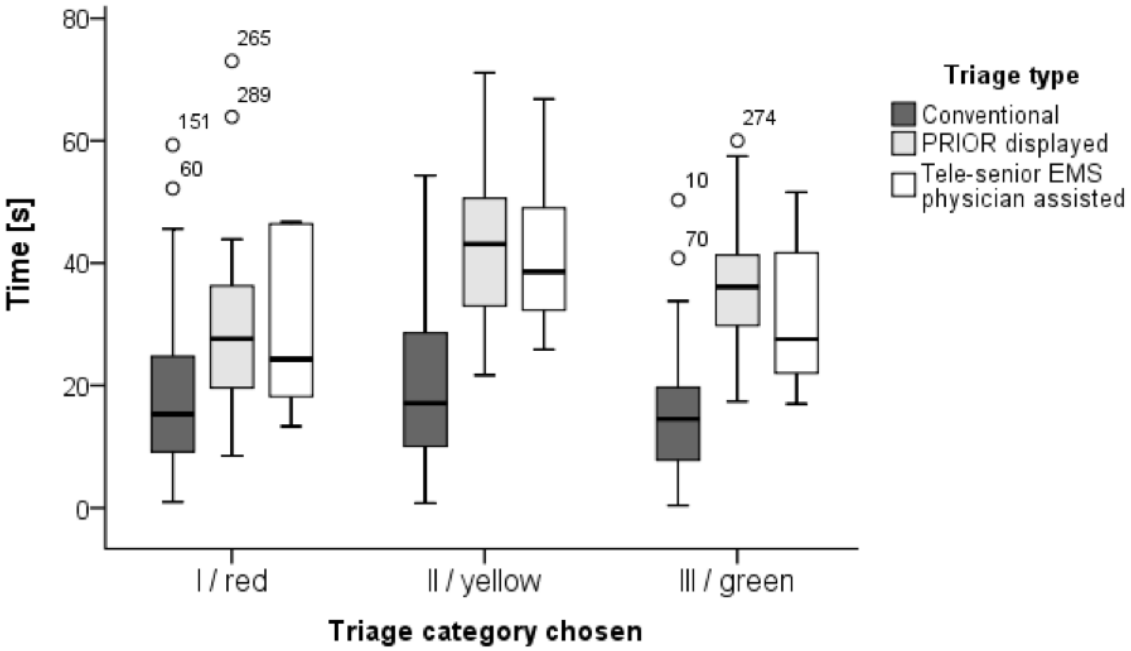
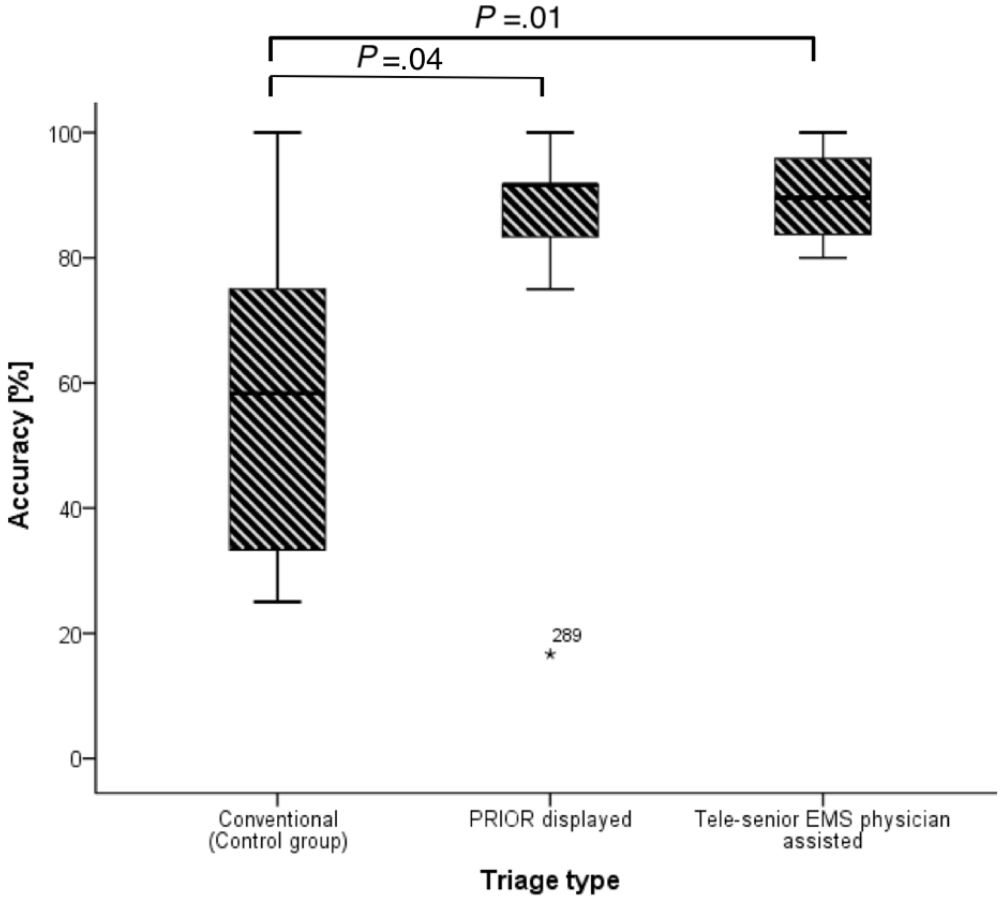


Figure 5. Accuracy (CI) of the chosen triage results per triage group. Numbers denote outliers. PRIOR: Primary Ranking for Initial Orientation in Rescue service; EMS: emergency medical service.



Discussion

Principal Findings

In a simulated MCI scenario, Smart Glasses apps for displaying a triage algorithm or telemedical contact with an experienced emergency physician were applied to support the triage conducted by paramedics compared with a conventional control group. While in one test group, the PRIOR triage algorithm was used to ensure standardized decision support, a telemedical connection to a senior EMS physician, who could also follow a triage algorithm or could deviate from it, was provided in the second test group. Although triage took almost 2 times as long in each test group when compared with the control group, the overall quality of triage—as measured by its accuracy—was markedly increased.

The duration of a conventional triage with 16.6 seconds is consistent with similar studies. It was noticeable that many of the participants did not apply the previously learned PRIOR, but often decided spontaneously; this was also reflected in the overall bad triage quality with a huge interindividual range. Only 58% of the triage corresponded to the correct results previously determined by 3 independent, experienced emergency physicians. Thus, the accuracy of triage remains under the experience of other studies that also investigated the use of different triage algorithms [16,17].

In the observation of subjects, many displayed distinct nervousness, which, in turn, confirmed the realism of the simulation. Many went into the so-called “tunnel vision,” and the triages were wrong. For the most part, a higher priority category was chosen. While this phenomenon is known as a potential source of error for the PRIOR algorithm, this extent indicates that the algorithm has rarely been correctly applied. Although the control group was explicitly instructed to use the algorithm, none of the subjects used the handed-out pocket card, which is also available for real disaster use in many EMS units. Numerous studies showed that checklists in medicine are not well accepted but regularly used in other safety-related professions, for example, in aviation [18,19].

Finally, those subjects who viewed the PRIOR triage algorithm on the Smart Glasses app showed markedly better triage quality, which was significantly increased with 92% matches compared with 58% for the control group. All participants in this study group were introduced to use the app for algorithm display for every triage. By working through the decision tree, the subjects were forced to use the PRIOR until the triage result was achieved. Owing to the direct display of the algorithm in front of the eyes compared with the pocket card, this tool was really used. However, this resulted in a longer duration of triage, which almost doubled compared with the control group.

In this context, it is important to note that this group has digitally recorded the triage results compared with a manual tally maintained by the control group; this resulted in many advantages with decisive importance in the deployment tactics. In addition, it allowed the digital results to be retrieved from anywhere and could, in future, contribute to the early knowledge of human and material needs; this is particularly important in a disaster scenario to better plan for rare resources such as hospital capacity and ambulances. However, an internet connection is not required to display the PRIOR algorithm in Smart Glasses but required for the digital acquisition of the triage results.

The telemedical assistance in triage was generated by a live video streaming through the camera of the Smart Glasses; this showed a particularly good triage quality of 90% and thus, a highly significant increase compared with the control group. This was certainly owing to both, the four eyes principle and the presence of a doctor in the screening process. However, this contradicts the statement of the paramedic in the questionnaire who considered a doctor in the triage pointless. The much better quality of triage is certainly also attributed to the high level of compliance with the guidelines for the processing of the screening algorithm, as well as other procedural instructions; this phenomenon is also indicated by other studies of individual medical treatments by the tele-EMS physician in Aachen Germany [20].

The duration of the tele-senior EMS physician-assisted triage was markedly prolonged, but a digital acquisition of the triage results was achieved, unlike the control group. In addition, the tele-physician was able to collect and document more information in addition to the triage category, such as the patient name and first diagnoses. The longer time required for classification into the category III of the slightly injured is a well-known phenomenon in the PRIOR algorithm, as well as a high rate of overtriage [2]. The time lag from technical support was the lowest in triage category I (Figure 4). As this category identifies severely injured patients with immediate treatment priority, this is an important finding.

The reason why no 100% accuracy in triage has been achieved, however, can best be explained by the different estimation of qualitative characteristics. Thus, it must be assessed whether there is a respiratory or circulatory disorder, without this being described in more detail. In addition, the group triage algorithm display on the data glasses did not achieve full accuracy. In addition to the above reasons, errors in the menu navigation of the data glasses and thus in the course of the algorithm can have an influence here.

In 2015, a feasibility study was carried out on the app of modern telemedicine in a disaster to triage, in which only 2 patients were triaged with telemedicine [21]; the authors found no marked differences in the quality of the triage, but in the duration. In another study, the use of optical head-mounted displays in disaster missions was mentioned as beneficial, without direct comparisons to a control group [22]. Another triage algorithm was tested on Google Glass during a full-scale exercise to perform visually guided augmented-reality Simple Triage and Rapid Treatment triage and identify casualties and collect georeferenced notes, photos, and videos to be incorporated into the debriefing [23]. However, this study demonstrated for the first time the controlled randomized comparison between conventional triage, the display of triage algorithms as augmented reality, and telemedically assisted triage.

The increased quality of triage by using the Smart Glasses was also reflected in the questioning of subjects. In this study, the majority of subjects who underwent triage with the technical support provided by the Smart Glasses described an increased sense of security; this can be an important factor for the emergency services in the stressful and unfamiliar situation of a disaster case.

Sufficient usability of the Smart Glasses was confirmed under realistic conditions. Operations using the optical touchpad for menu control of the Smart Glasses was also done while wearing protective gloves, which this led to no considerable problems, although only a very short briefing was given when handling the Smart Glasses. In addition, the simulation was completed by subjects using their own personal protective equipment, and sufficient compatibility was confirmed in the questionnaire.

Limitations

The different size of the study groups resulted from the fact that the control group was also part of a parallel observational study. The size of the 2 study groups with technical support from the data glasses (with both the display of the algorithm and telemedical contact) was originally designed with n=10. However, owing to the given timeslots, not all paramedics could participate; this fact is a limitation that should be considered in future studies.

Another limitation, in addition to the low battery capacity, which could be extended by connecting a battery pack (Figure 2), there is a lack of compatibility with personal glasses. For eyeglass wearers to be able to use the Smart Glasses, personal glasses would have to be adapted; this would involve considerable costs. In noneyeglass wearers, however, the current Smart Glasses also serve as protective eyewear. In addition, the local Wi-Fi connection led to problems. In 10 triages, no adequate connection could be achieved owing to structural obstacles; these were excluded from the study results. A mobile connection with sufficient network coverage and increased battery capacity would certainly make sense here. Unfortunately, such solutions are currently not available on the market.

Conclusions

In summary, technically assisted triage shows markedly better quality than traditional methods. Smart Glasses have proven to be a useful tool in disaster medicine; they allow EMS responders to continue working with both hands while information is displayed on the monitor and data are collected through the integrated camera. The delay of the triage seems acceptable and in view of the digital coverage of the triage quality. Further developments to the system, as well as use in routine operations, will certainly shorten the duration of triage markedly. Assuming that both augmented and virtual reality will gain in importance in the coming years in both work and leisure, future users will then use it much more quickly and intuitively.

This study has shown that research on screening assistance procedures is still required to achieve sufficient quality of triage, which can be critical and contribute to the targeted and prioritized treatment and transport of patients in a disaster situation. High-tech can thus also support special challenges in the event of a disaster. These missions are rare and therefore lacking in routine with all emergency medical professionals. It is important to exploit all potentials of modern technologies in such situations and integrate and use augmented reality and telemedicine in existing civil defense structures.

Acknowledgments

Funding for the AUDIME project was provided by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung).

Authors' Contributions

AF, MO, NH, and MC conceived the study, designed the trial, and obtained research funding. AF, NH, RR, and MC supervised the conduct of the trial and data collection. AF and MC managed the data, including quality control. SKB and RR provided statistical advice on study design. AF and MC analyzed the data. AF drafted the manuscript, and all authors contributed substantially to its revision. AF takes responsibility for the paper as a whole.

Conflicts of Interest

MO and MC are employees of Docs in Clouds GmbH, Aachen, Germany; MC and RR are shareholders of Docs in Clouds GmbH, Aachen, Germany.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 598KB - jmir_v21i1e11939_app1.pdf]

References

1. Clarkson L, Williams M. Triage, Mass Casualty. In: StatPearls [Internet]. Treasure Island, FL: StatPearls Publishing; 2017.
2. Heller AR, Salvador N, Frank M, Schiffner J, Kipke R, Kleber C. Diagnostic precision of triage algorithms for mass casualty incidents. English version. Anaesthesist 2017 Aug 10:1-10. [doi: 10.1007/s00101-017-0352-y] [Medline: 28798972]

3. Bubser F, Callies A, Schreiber J. PRIOR: Vorsichtungssystem für Rettungsassistenten und Notfallsanitäter. Rettungsdienst 2014;37(8):730-734.
4. Ellebrecht N, Latasch L. Vorsichtung durch Rettungsassistenten auf der Großübung SOGRO MANV 500. Notfall Rettungsmed 2011 Jul 31;15(1):58-64. [doi: 10.1007/s10049-011-1477-1]
5. Hoff JJ, Carroll G, Hong R. Presence of undertriage and overtriage in simple triage and rapid treatment. Am J Disaster Med 2017;12(3):147-154. [doi: 10.5055/ajdm.2017.0268] [Medline: 29270957]
6. Kilner TM, Brace SJ, Cooke MW, Stallard N, Bleetman A, Perkins GD. In 'big bang' major incidents do triage tools accurately predict clinical priority?: a systematic review of the literature. Injury 2011 May;42(5):460-468. [doi: 10.1016/j.injury.2010.11.005] [Medline: 21130438]
7. Dadoun E, Khalil E, Bank I. Invoking the “expectant” triage category: Can we make the paradigm shift? Am J Disaster Med 2017;12(3):167-172. [doi: 10.5055/ajdm.2017.0270] [Medline: 29270959]
8. Jenkins JL, McCarthy ML, Sauer LM, Green GB, Stuart S, Thomas TL, et al. Mass-casualty triage: time for an evidence-based approach. Prehosp Disaster Med 2008;23(1):3-8. [Medline: 18491654]
9. Nakao H, Ukai I, Kotani J. A review of the history of the origin of triage from a disaster medicine perspective. Acute Med Surg 2017 Dec;4(4):379-384 [FREE Full text] [doi: 10.1002/ams2.293] [Medline: 29123897]
10. Paulus A, Czaplik M, Hirsch F, Meisen P, Meisen T, Jeschke S. AUDIME: Augmented Disaster Medicine. In: Jeschke S, Isenhardt I, Hees F, Henning K, editors. Automation, Communication And Cybernetics In Science And Engineering. Berlin: Springer Verlag; 2018:637-644.
11. Brokmann JC, Rossaint R, Bergrath S, Valentin B, Beckers SK, Hirsch F, et al. Potential and effectiveness of a telemedical rescue assistance system. Prospective observational study on implementation in emergency medicine. Anaesthesist 2015 Jun;64(6):438-445. [doi: 10.1007/s00101-015-0039-1] [Medline: 26036316]
12. Brokmann JC, Rossaint R, Hirsch F, Beckers SK, Czaplik M, Chowanetz M, et al. Analgesia by telemedically supported paramedics compared with physician-administered analgesia: A prospective, interventional, multicentre trial. Eur J Pain 2016 Dec;20(7):1176-1184. [doi: 10.1002/ejp.843] [Medline: 26914284]
13. Brokmann JC, Conrad C, Rossaint R, Bergrath S, Beckers SK, Tamm M, et al. Treatment of Acute Coronary Syndrome by Telemedically Supported Paramedics Compared With Physician-Based Treatment: A Prospective, Interventional, Multicenter Trial. J Med Internet Res 2016 Dec 01;18(12):e314 [FREE Full text] [doi: 10.2196/jmir.6358] [Medline: 27908843]
14. Czaplik M, Bergrath S, Rossaint R, Thelen S, Brodziak T, Valentin B, et al. Employment of telemedicine in emergency medicine. Clinical requirement analysis, system development and first test results. Methods Inf Med 2014;53(2):99-107. [doi: 10.3414/ME13-01-0022] [Medline: 24477815]
15. Felzen M, Brokmann JC, Beckers SK, Czaplik M, Hirsch F, Tamm M, et al. Improved technical performance of a multifunctional prehospital telemedicine system between the research phase and the routine use phase - an observational study. J Telemed Telecare 2017 Apr;23(3):402-409. [doi: 10.1177/1357633X16644115] [Medline: 27080747]
16. Cross KP, Cicero MX. Head-to-head comparison of disaster triage methods in pediatric, adult, and geriatric patients. Ann Emerg Med 2013 Jun;61(6):668-676.e7. [doi: 10.1016/j.annemergmed.2012.12.023] [Medline: 23465555]
17. Mistry B, Stewart De Ramirez S, Kelen G, Schmitz PSK, Balhara KS, Levin S, et al. Accuracy and Reliability of Emergency Department Triage Using the Emergency Severity Index: An International Multicenter Assessment. Ann Emerg Med 2018 Dec;71(5):581-587.e3. [doi: 10.1016/j.annemergmed.2017.09.036] [Medline: 29174836]
18. Jain D, Sharma R, Reddy S. WHO safe surgery checklist: Barriers to universal acceptance. J Anaesthesiol Clin Pharmacol 2018;34(1):7-10 [FREE Full text] [doi: 10.4103/joacp.JOACP_307_16] [Medline: 29643615]
19. Thomassen Ø, Espeland A, Søfteland E, Lossius HM, Heltne JK, Brattebø G. Implementation of checklists in health care; learning from high-reliability organisations. Scand J Trauma Resusc Emerg Med 2011 Oct 03;19:53 [FREE Full text] [doi: 10.1186/1757-7241-19-53] [Medline: 21967747]
20. Bergrath S, Reich A, Rossaint R, Rörtgen D, Gerber J, Fischermann H, et al. Feasibility of prehospital teleconsultation in acute stroke--a pilot study in clinical routine. PLoS One 2012;7(5):e36796 [FREE Full text] [doi: 10.1371/journal.pone.0036796] [Medline: 22629331]
21. 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] [Medline: 25571779]
22. Berndt H, Mentler T, Herczeg M. Optical Head-Mounted Displays in Mass Casualty Incidents: Keeping an Eye on Patients and Hazardous Materials. Int J Inf Syst Crisis Response Manag 2015;30(1):1-15. [doi: 10.4018/IJISCRAM.2015070101]
23. Carenzo L, Barra FL, Ingrassia PL, Colombo D, Costa A, Della Corte F. Disaster medicine through Google Glass. Eur J Emerg Med 2015 Jun;22(3):222-225. [doi: 10.1097/MEJ.0000000000000229] [Medline: 25460812]

Abbreviations

AUDIME: Augmented Disaster Medicine
EMS: emergency medical service

IQR: interquartile range
MCI: mass casualty incident
PRIOR: Primary Ranking for Initial Orientation in Rescue service

Edited by G Eysenbach; submitted 14.08.18; peer-reviewed by K Goniewicz, J Salisbury; comments to author 19.09.18; revised version received 06.10.18; accepted 06.10.18; published 03.01.19.

Please cite as:
Follmann A, Ohligs M, Hochhausen N, Beckers SK, Rossaint R, Czaplik M
Technical Support by Smart Glasses During a Mass Casualty Incident: A Randomized Controlled Simulation Trial on Technically Assisted Triage and Telemedical App Use in Disaster Medicine
J Med Internet Res 2019;21(1):e11939
URL: https://www.jmir.org/2019/1/e11939/
doi:10.2196/11939
PMID:30609988

Original Paper

Informing Adults With Back Pain About Placebo Effects: Randomized Controlled Evaluation of a New Website With Potential to Improve Informed Consent in Clinical Research

Felicity L Bishop[1], PhD; Maddy Greville-Harris[1], PhD; Jennifer Bostock[2], MA; Amy Din[3], MSc; Cynthia A Graham[1], PhD; George Lewith[4], FRCP; Christina Liossi[1], DPsych; Tim O'Riordan[5], MSc; Peter White[3], PhD; Lucy Yardley[1,6], PhD

[1]Department of Psychology, University of Southampton, Southampton, United Kingdom

[2]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[3]Centre for Innovation & Leadership in Health Sciences, School of Health Sciences, University of Southampton, Southampton, United Kingdom

[4]Primary Care and Population Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[5]Zemedia, Southampton, United Kingdom

[6]School of Psychological Science, Faculty of Life Sciences, University of Bristol, Bristol, United Kingdom

Corresponding Author:
Felicity L Bishop, PhD
Department of Psychology
University of Southampton
Building 44, Highfield Campus
University Road
Southampton,
United Kingdom
Phone: 44 2380599020
Email: f.l.bishop@southampton.ac.uk

Abstract

Background: Placebo effects and their underpinning mechanisms are increasingly well understood. However, this is poorly communicated to participants in placebo-controlled trials. For valid informed consent, participants should be informed about the potential benefits and risks of participating in placebo-controlled trials. Existing information leaflets often fail to describe the potential benefits and adverse effects associated with placebo allocation. This study tested the effects of a new website designed to inform patients about placebo effects (The Power of Placebos, PoP). PoP was designed using qualitative methods in combination with theory- and evidence-based approaches to ensure it was engaging, informative, and addressed patients' concerns.

Objective: This study aimed to test the effects of PoP, compared with a control website, on people's knowledge about placebo and the ability to make an informed choice about taking part in a placebo-controlled trial.

Methods: A total of 350 adults with back pain recruited from 26 general practices in Southern England participated in this Web-based study. Participants were randomly assigned to PoP (which presented scientifically accurate information about placebo effects in an engaging way) or a control website (based on existing information leaflets from UK trials). Participants self-completed Web-based pre- and postintervention questionnaire measures of knowledge about placebo effects and preintervention questionnaire measures of attitudes toward and intentions to participate in a placebo-controlled trial. The 2 primary outcomes were (1) knowledge and (2) informed choice to take part in a placebo-controlled trial (computed from knowledge, attitudes, and intentions).

Results: After viewing PoP, participants had significantly greater knowledge about placebos (mean 8.28 [SD 1.76]; n=158) than participants who viewed the control (mean 5.60 [SD 2.24]; n=174; $F_{1,329}$=173.821; P<.001; η^2=.346). Participants who viewed PoP were 3.16 times more likely than those who viewed the control to make an informed choice about placebos (χ^2_1=36.5; P<.001).

Conclusions: In a sample of adults with back pain, PoP increased knowledge and rates of informed choice about placebos compared with a control website. PoP could be used to improve knowledge about placebo effects in back pain. After essential further development and testing in clinical trial settings, it could support informed consent in placebo-controlled trials.

(J Med Internet Res 2019;21(1):e9955) doi:10.2196/jmir.9955

KEYWORDS

placebos; placebo effects; informed consent; research ethics; health knowledge, attitudes, practice; internet

Introduction

Background

Placebo-controlled trials remain the gold standard for establishing the efficacy of new pharmacological and other interventions and are often used in pain medicine. According to the Declaration of Helsinki, investigators should obtain a priori voluntary informed consent from participants after informing them about "the anticipated benefits and potential risks of the study and the discomfort it may entail" (item 26, [1]). Although participant information leaflets typically focus on the benefits and risks of the new intervention being trialed [2], placebos themselves can trigger both beneficial and adverse effects in a range of conditions, particularly pain [3-7]. Patients in a trial who receive the placebo may thus experience placebo and/or nocebo effects but are unlikely to have been informed about them in advance. Arguably, it should be standard practice to inform patients in placebo-controlled trials about the placebo and its possible benefits and adverse effects [8].

Current information provision about placebo interventions in clinical trials is likely to be inadequate and thus jeopardizes the validity of informed consent. Qualitative studies and surveys suggest that members of the public and trial participants often have misunderstandings and partial knowledge about placebos and their effects [9,10]. For example, trial participants often believe that placebo effects are fake or illusory and that people who respond to placebos are gullible or foolish [10-12]. Such beliefs are inconsistent with the scientific literature on placebo effects and may: deter people from volunteering for trials [13]; contribute to patient anxiety about placebo effects [11]; and/or make it difficult for participants to make sense of a personal placebo response when debriefed to placebo allocation at the end of a trial [14,15]. Furthermore, only a small minority of people understand that placebos can have adverse effects; this has been found consistently across surveys of trial participants [9] and patients [16-18]. These beliefs are also unlikely to be challenged by existing patient information leaflets, which barely mention the placebo; a content analysis of participant information leaflets from major UK-based placebo-controlled trials found that only 1 of the 45 leaflets studied explicitly stated that patients might experience beneficial effects from the placebo, only 4 leaflets explicitly stated that patients might experience adverse effects from the placebo, and 8 leaflets explicitly stated that the placebo treatment was either undesirable or ineffective [2]. This paucity of information about placebos is probably widespread, as studies of patient information leaflets have reported similar findings, for example, in Finland [19], Spain [20], and internationally [21].

Preliminary evidence suggests it is possible to improve people's understandings of placebos [22,23], but effects in well-defined clinical populations and on key outcomes such as informed choice have not been evaluated. Web-based resources could usefully augment or improve existing paper-based information leaflet because websites (1) are increasingly popular with health consumers [24,25]; (2) easily incorporate interactive features [26], which can enhance engagement and effective education [27]; (3) are easily and cheaply disseminated for widespread access [26]; and (4) can be readily adapted and/or tailored for use in different clinical trials [26].

Objectives

A Web-based experiment was conducted to compare a newly developed, scientifically accurate, and engaging website about placebos—The Power of Placebos (PoP) [28]—with a control website based on existing patient information about placebos. PoP was designed for and tested in adults with back pain because placebo analgesia is well documented in this population [3,5,29,30]. It was important to focus on 1 condition as the nature of and mechanisms underpinning placebo effects differ by disease [31]. Back pain is a major public health concern [32] as it is the leading cause of disability in most countries [33]. The most effective therapies appear to involve exercise and education [34], although 1 study has found improvements in back pain after open-label placebo treatment [35]. The hypotheses were that PoP would increase knowledge about placebos, enable more patients to make an informed choice about having placebos, and encourage people to believe that placebos can have credible analgesic and adverse effects.

Methods

Design

This Web-based study tested the effects of a newly developed interactive website (PoP) on patients' knowledge, informed choice, and beliefs about placebos. Participants were randomized automatically by the overarching study website to view either PoP or the control website. They also saw a separate website about acupuncture (again either a newly developed one or a control), as this study was part of a larger 2 × 2 factorial trial (for details, see Multimedia Appendix 1). As there were no effects of the acupuncture website manipulation on patients' knowledge, informed choice, and beliefs about placebos, this study reports the PoP versus control comparison for the placebo website only. The equivalent comparison for the acupuncture website is being reported separately.

Furthermore, 2 volunteers, with personal experience of back pain and clinical research, acted as patient advisors in this study. Research has shown that such "patient public involvement" in research can enhance the design and conduct of studies [36]. Both our volunteers participated in team meetings, which involved discussions and decision making around study design, procedures, and conduct. The patient advisors also helped to finalize recruitment materials and contributed to website design, for example, by providing feedback on prototypes of PoP. Of the 2 advisors, 1 advisor also chose to contribute to the process of writing up for publication.

Ethical approval was sought and obtained from the University of Southampton (Ethics ID: 12323) and the NHS NRES (National Health Service National Research Ethics Service) Committee East of England-Hatfield (Research Ethics Committee reference: 14/EE/1176).

Participants and Recruitment

Adults aged 18 years and older with a recent history of back pain (within 3 years) as documented by their general practitioner (GP) were recruited via 26 general practices in South West England between December 2014 and March 2015. General practice staff conducted database searches and mailed study invitation packs to eligible patients. Invitation packs contained a cover letter and information sheet including the study website address. Participants needed to be computer literate to self-enroll in the study and complete it independently (technical support was available from the researchers by telephone or email if needed). People with needle phobia (important for the acupuncture websites) or unable to complete questionnaires in English were excluded. Figure 1 shows the flow of participants through the study.

An a priori power calculation was conducted using G*Power (Heinrich Heine Universitat Dusseldorf) [37]. Assuming an effect size f=0.15 (based on unpublished pilot data), power 0.8, and alpha .05 for a factorial analysis of variance (ANOVA) required a total of 351 participants, and assuming 5% dropout required 369 patients to be randomized.

Interventions

Overall, 2 websites about placebo effects were developed for this study: a person-based website and a control website. Both websites were explicitly labeled as developed at the University of Southampton and included identical brief biographies of the research team. Participants were not told which website they were viewing. Both websites presented information about placebo effects and were created using LifeGuide software, a package developed for researchers to design, build, and trial Web-based interventions [38]. Guidance for developing patient-focused health information was followed to ensure both websites were understandable, readable, and accessible [39]. The person-based website was longer than the control website (10 pages vs 5 pages). Table 1 compares the content and formats of each website. No changes to the websites were made during this study.

Figure 1. Flow of participants through the study. GP: general practitioner; PoP: The Power of Placebos.

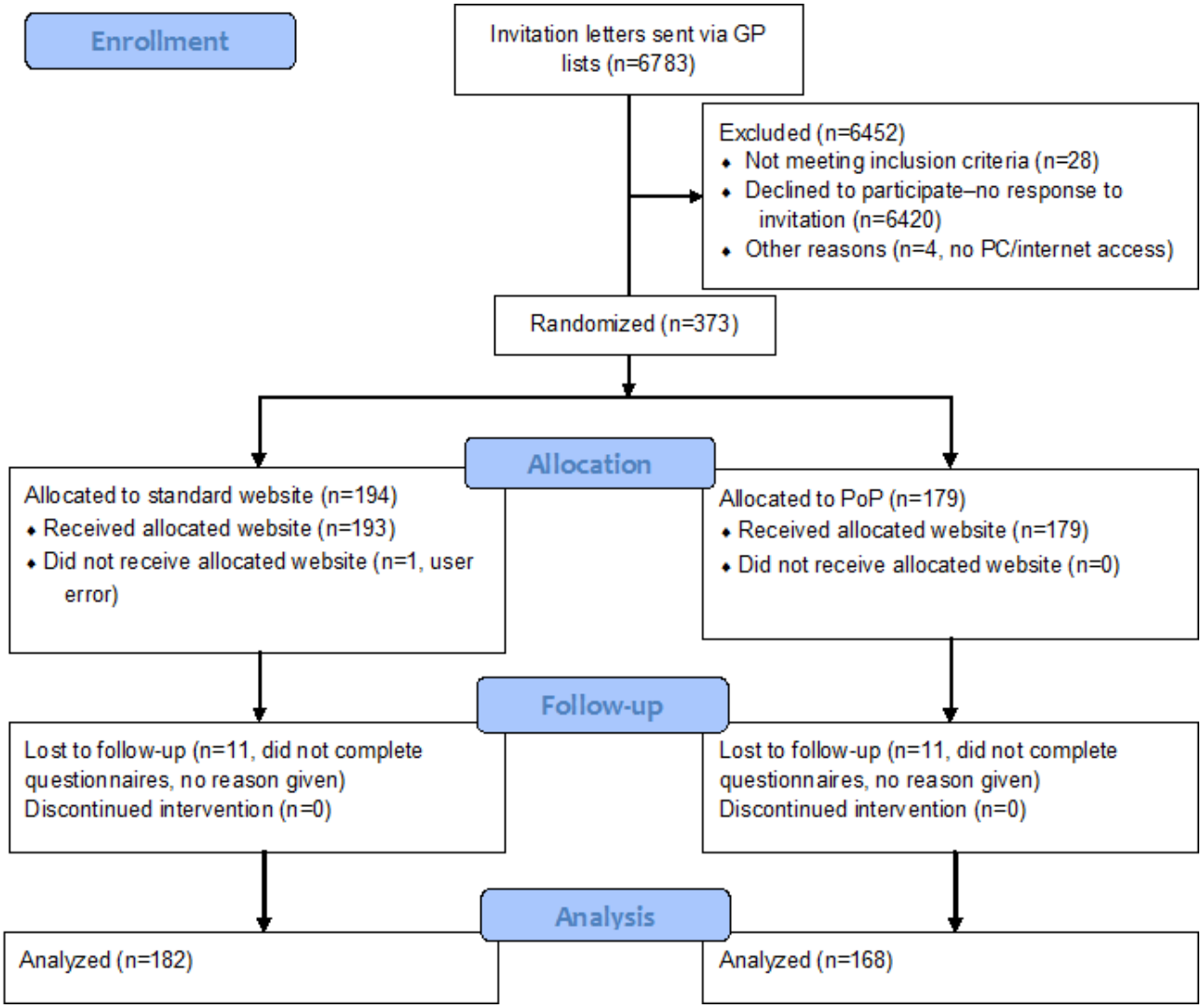


Table 1. Summary comparison of The Power of Placebos (PoP) and control websites.

Content and format of websites	PoP	Control
Contents		
Defining placebos	✓[a]	✓
Potential benefits of placebo	✓	✓
Potential adverse effects of placebo	✓	X[b]
Patients' experiences of placebos	✓	X
Common concerns about placebos	✓	X
Debunking myths that placebo responders are malingerers or gullible	✓	X
Mechanisms underpinning placebo effects	✓	X
Placebos in placebo-controlled trials	✓	✓
Placebos in clinical practice	✓	X
Formats		
Text	✓	✓
Images	✓	✓
Film	✓	X
Audio clips	✓	X

[a]Tick indicates feature is present in the website.

[b]Cross indicates feature is absent in the website.

The Power of Placebos Website

PoP was developed using a rigorous approach derived from person-based intervention development, which incorporates evidence and theory [40-42]. This entailed extensive intervention planning, drawing on existing evidence and theory about placebo effects and how to change patients' health beliefs and behaviors, and conducting qualitative think-aloud interviews to develop and refine the website. The content was based on published scientific evidence [3,4,31,43] and targeted theoretically informed constructs (primarily knowledge but also attitudes, subjective norms, and perceived behavioral control, derived from the Theory of Planned Behavior [44]). For a full description of the development process and the resulting website, refer to the study by Greville-Harris et al [28]. Figure 2 shows an example page, and screenshots showing additional pages are available in Multimedia Appendix 2. A completed TIDieR (template for intervention description and replication) checklist in Multimedia Appendix 3 presents full information about the description of PoP.

The Control

The control website was based on the limited information about placebos that is included in some UK patient information leaflets for placebo-controlled clinical trials. Consistent with common practice among those leaflets that provide any information about placebos [2], it provided only minimal detail about the possible effects of placebos and did not explain their mechanisms of action. Figure 3 shows an example page. Screenshots of the entire control website are available in Multimedia Appendix 4.

Measures

Participants' Characteristics

Items selected from the recommended minimum dataset for back pain [45] were used to assess clinical characteristics (pain duration, pain frequency, pain intensity, pain catastrophizing, and pain spread to legs), pain-related legal claims, disability benefits or compensation (all single items), and pain functioning and interference (4-item scales with excellent internal consistency in this sample; Cronbach alphas .96 and .92, respectively). Single items assessed ethnicity, age, gender, and education level.

Primary Outcomes

Primary outcomes were knowledge and informed choice about placebos. Informed choice was selected as an analogy to informed consent because the website was being tested outside the context of a placebo-controlled trial. Informed choice is a composite measure based on knowledge, attitudes, and intentions to try placebo, the components of which were measured after participants had finished viewing the website.

Figure 2. Screenshot from the Power of Placebos website: How do placebos work?

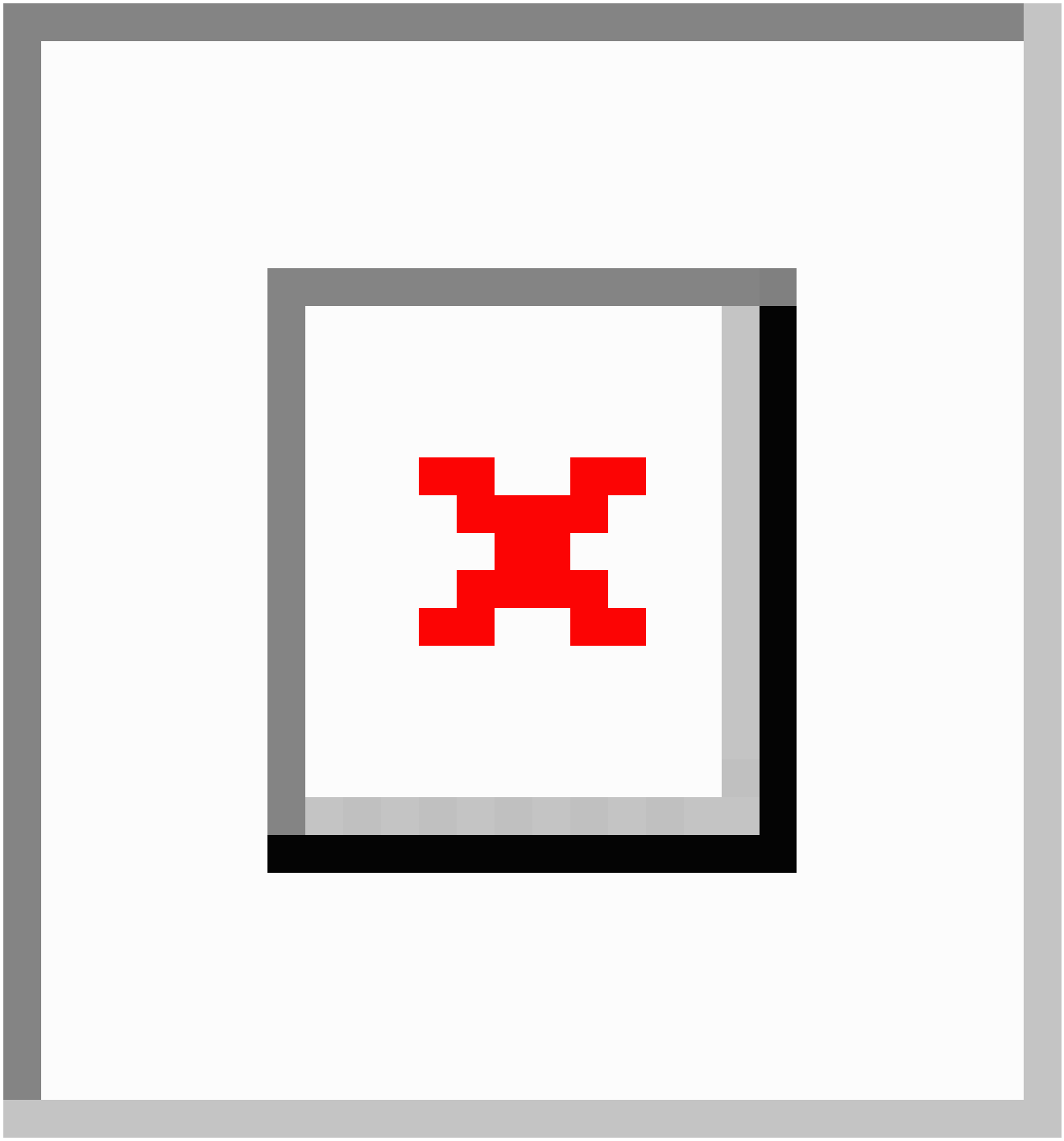
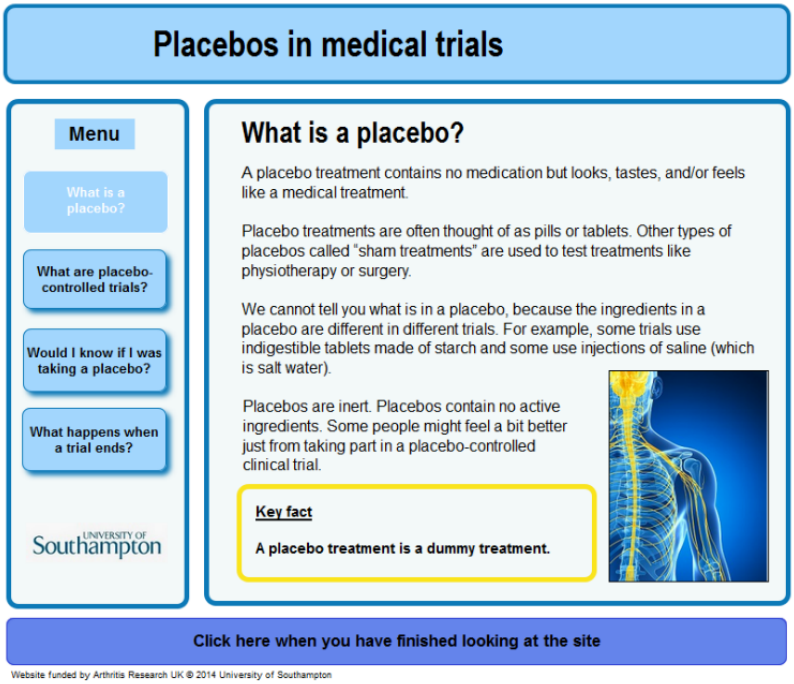


Figure 3. Screenshot from control webpage.



Placebo knowledge was assessed using a 10-item knowledge quiz, comprising true-false questions developed and validated in preparation for this study. These 10 items were selected from a larger pool of 15 items pilot tested in a Web-based study with a community-based sample of 210 adults with a history of back pain [18]. The 10 items most commonly answered incorrectly by the community-based sample were chosen for this study (eg, "A placebo pill can have side effects" and "Placebo pills can help to treat pain conditions"—both are true). The knowledge score is the total number of items answered correctly. The quiz was completed before and after viewing the websites. The pre- and postadministrations had acceptable internal consistency in this sample (Cronbach alphas .66 and .79, respectively).

Making an informed choice can be understood as choosing to act in a way that is based on one's knowledge and one's values [46-48]. To make an informed choice, a person needs to have an accurate understanding of the options available, have formed an opinion about the options based on their personal values, and make a decision (or otherwise act in a way) that is consistent with their knowledge and values [46-48]. Although it can be argued that knowledge alone is sufficient for informed choice, attitudes are incorporated in our chosen definition as an attempt to reflect the role of personal values in making health decisions that are optimized for the individual concerned. If decisions were based on knowledge alone, people could still make decisions that are inconsistent with their personal values and so could be considered sub-optimal for them as individuals. By choosing a definition of informed choice that incorporates attitudes, we are better able to model the role of personal values in decision making. From this perspective, an informed choice to have a placebo requires knowledge about the possible beneficial and adverse effects of placebos, a positive attitude to placebos, and a decision to try a placebo. An informed choice not to have a placebo requires knowledge about the possible beneficial and adverse effects of placebos, a negative attitude to placebos, and a decision not to try a placebo.

In this study, knowledge was measured using the placebo knowledge quiz. Attitudes were measured using 4 items derived following the Theory of Planned Behavior guidelines for measuring attitudes [49], for example, "having a placebo treatment would be good." Behavioral intentions were used as a proxy for behavior and were measured using 3 items derived following the same guidelines [49], for example, "if given the opportunity, I intend to have a placebo treatment." Items assessing attitudes and intentions were measured on 7-point Likert-type scales labeled strongly disagree to strongly agree, and scores across constituent items were averaged. The scales had good internal consistency (Cronbach alphas .97 and .87, respectively).

Participants were categorized as making an informed choice or not based on their attitude (split by scale midpoint to produce 2 groups with positive and negative attitudes), intention to have treatment (split by scale midpoint to produce 2 groups with positive and negative intentions), and their knowledge score (based on median split to produce 2 groups with high and low knowledge). Participants were categorized as making an informed choice if they scored above the median on knowledge and either (a) above the scale midpoint on both attitudes and intentions or (b) below the scale midpoint on both attitudes and intentions. All other participants were categorized as not making an informed choice. Although there are drawbacks of computing composite measures in this way (see Discussion), this approach enabled us to follow published definitions of informed choice to create a primary outcome measure with good face validity and is consistent with previous studies of informed choice in other settings [46-48].

Secondary Outcomes

Secondary outcomes were participants' beliefs about placebos and willingness to try a placebo. These were completed after participants had finished viewing the website.

Overall, 4 dimensions of beliefs about placebos were measured using the previously validated 4-item subscales of the Low Back Pain Treatment Beliefs Questionnaire [50]: concerns (eg, "I have concerns about having placebos for my back pain"), individual fit (eg, "I am confident placebos would be a suitable treatment for my back pain"), expectancy (eg, "Placebos can work well for people with back pain"), and credibility (eg, "Generally, placebos are a believable therapy for back pain"). The Low Back Pain Treatment Beliefs Questionnaire was validated in a Web-based study [50]. All items used 5-point Likert-type response scales labeled strongly disagree to strongly agree. All subscales had good internal consistency in this sample (Cronbach alpha for concerns=.72, individual fit=.91, expectancy=.94, and credibility=.84).

A single yes or no item asked whether participants would be "willing to have placebo treatment." As participants could take breaks from viewing the website and return later, an item was also included to assess whether participants had "looked up additional information about placebo during breaks from the study."

Procedure

On accessing the study website at a time and location of their choice (eg, at home), not in the presence of a researcher, participants viewed the information sheet and indicated consent to take part by clicking a button (see Multimedia Appendix 5 for information sheet). Participants were then asked 3 mandatory screening questions assessing age, current or recent back pain, and needle phobia. Those not meeting the associated inclusion criteria were directed to an exit page. Those passing the screening questions entered their email address and created a password for the website, which allowed participants to take breaks when desired. Participants then completed the 10-item placebo knowledge quiz to assess baseline knowledge before being presented with the series of 2 websites (placebo and acupuncture) appropriate to their randomization group. Participants could take breaks, log out and return to the study later, and stop viewing each website whenever they wanted ("click here when you have finished looking at the information" was available on every page). Directly after viewing the websites, participants completed the primary and secondary outcome measures and measures of sociodemographic and clinical characteristics. All measures were completed on the study website. Finally, participants were directed to a debriefing page with further information about the study and links to other resources. Everyone who completed the study was emailed a £10 online shopping voucher.

Statistical Analysis

Data were downloaded from LifeGuide and imported into IBM SPSS Statistics v22 for analysis. The proportion of missing data was small (<5% on any 1 variable) but was not missing completely at random (MCAR; Little MCAR test: χ^2_{7845}=8279.7; P<.001), suggesting imputation might be inappropriate but unlikely to alter the results [51]. All analyses were repeated excluding missing data and then imputing missing values with the expectation-maximization algorithm. The results were the same, and the reported analyses used all available data with no imputation.

Pearson's chi-square test compared the proportions of people (a) making an informed choice and (b) willing to have a placebo, between PoP and the control website. ANOVAs tested the effects of PoP compared with control on postintervention measures of knowledge and treatment perceptions. Analyses of covariance were then computed adjusting for baseline knowledge [52], which was the only baseline variable to differ between the groups.

Results

Participants' Characteristics

The final sample comprised 350 adults, of whom the majority were female, white British, and educated up to at least 18 years (see Table 2). The average age of participants was 47.88 years (SD 15.8; control group mean age 48.80 years [SD 15.9]; PoP group mean age 46.88 years [SD 15.6]). Almost all participants (308/350, 88.0%) reported having back pain in the last week, and on average, they reported pain interfering with daily activities between *a little bit* and *somewhat* and being able to perform functional tasks such as chores or walking *with a little difficulty*. Participants' back pain was typically longstanding (157/350, 45.1% had onset over 5 years ago), affected them daily or almost daily (133/350, 38.0%), and was of moderate intensity (mean 4.75 on a 10-point scale; control group mean 4.79 [SD 2.46]; PoP group mean 4.71 [SD 2.40]). Pain interference in the past week was also moderate (control group mean 2.58 [SD 1.29]; PoP group mean 2.58 [SD 1.25]; on a 5-point scale) as was current pain functioning (control group mean 1.98 [SD 0.96]; PoP group mean 2.06 [SD 1.02]; on a 5-point scale). Approximately, one-third (126/350, 36.0%) reported pain catastrophizing. Very few participants (2 per group) reported looking up additional information about placebo effects during the study. The groups did not differ significantly on any of these measures (all P values >.05), except baseline knowledge about placebo, on which the PoP group scored slightly higher (mean 6.21 [SD 2.14]) than the control group (mean 5.72 [SD 2.20]; t_{340}=−2.08; P=.04).

Knowledge, Attitudes, and Intentions

There was a significant main effect of website on placebo knowledge: after viewing the website, people who viewed PoP had higher knowledge about placebo effects than people who viewed the control website (Table 3). On average, people who viewed PoP also had less positive attitudes and more positive intentions toward placebo effects than people who viewed the control website, but these effects were much smaller than the effect on knowledge (Table 3).

Informed Choice

Table 4 shows how participants were classified as making an informed choice or not according to their knowledge, attitudes, and intentions. There was a significant association between

website and informed choice about placebo (χ^2_1=36.5; *P*<.001). Almost half of people (66/146, 45.2%) who viewed PoP made an informed choice about placebos compared with 14.3% (24/168) of people who viewed the control website. Thus, people who viewed PoP were 3.16 times more likely than those who viewed the control website to make an informed choice about placebos. The most common pattern of scores on components of informed choice was to have negative intentions of having placebos, *and* negative attitudes toward placebos, *and* low knowledge about placebos, a pattern displayed by 27.4% of all participants, 41.1% of participants who viewed the control website, and 11.6% of participants who viewed PoP.

Treatment Beliefs

After viewing the website, people who viewed PoP had significantly more positive beliefs that placebo treatment was a good fit for them, had significantly higher expectations of benefit from placebo, and perceived placebo treatment as significantly more credible than people who viewed the control website (Table 5). Level of concerns about placebo treatment did not differ between the groups.

Willingness to Try Placebo

There was a significant effect of website on willingness to try a placebo (χ^2_1=10.1; *P*=.001). More than half of participants (59.3%, 99/167) who had viewed PoP compared to 42.2% (76/180) of those who had viewed the control website said they would be willing to try a placebo. Thus, people who viewed PoP were 1.41 times more likely than those who viewed the control website to be willing to try a placebo.

Table 2. Participants' characteristics by group.

Characteristic and category	Frequency, n (%)		
	Whole sample (N=350)	Control website (n=182)	PoP[a] (n=168)
Sociodemographic characteristics			
Gender			
Female	197 (56.3)	102 (56.0)	95 (56.5)
Ethnicity			
White British	311 (88.9)	166 (91.2)	145 (86.3)
White (any other)	16 (4.6)	7 (3.8)	9 (5.4)
Asian or Asian British	4 (1.2)	1 (0.5)	3 (1.8)
Mixed ethnicity	2 (0.6)	0 (0.0)	2 (1.2)
Black or black British	2 (0.6)	1 (0.5)	0 (0.0)
Education			
Did not complete secondary school	19 (5.4)	13 (7.1)	6 (3.6)
Secondary school	89 (25.4)	48 (26.4)	41 (24.4)
Sixth form or college (aged 16-18 years)	106 (30.3)	61 (33.5)	45 (26.8)
Undergraduate study	98 (28.0)	43 (23.6)	55 (32.7)
Postgraduate study	35 (10.0)	16 (8.8)	19 (11.3)
Clinical characteristics			
Time since pain onset			
Up to 1 year	71 (20.4)	29 (15.9)	42 (25.0)
1 to 5 years	105 (30.2)	57 (31.3)	48 (28.6)
>5 years	157 (45.1)	83 (45.6)	74 (44.0)
Pain frequency in past 6 months			
Every day or nearly every day	133 (38.0)	70 (38.5)	63 (37.5)
More than half the days	85 (24.3)	40 (22.0)	45 (26.8)
Less than half the days	102 (29.1)	53 (29.1)	49 (29.2)
Disability or compensation benefits	16 (4.6)	7 (3.8)	9 (5.4)
Legal claim related to back	4 (1.1)	1 (0.5)	3 (1.8)
Pain spread to leg or legs in past 2 weeks	142 (40.6)	73 (40.1)	69 (41.1)
Pain catastrophizing	126 (36.0)	66 (36.3)	60 (35.7)
Previous participation in a placebo-controlled trial	6 (1.7)	4 (2.2)	2 (1.2)
Looked up additional information about placebos during the study	4 (1.1)	2 (1.1)	2 (1.2)

[a]PoP: The Power of Placebos.

Table 3. Postintervention knowledge, attitudes, and intentions toward placebos by group.

Measure	Control website		The Power of Placebos		Comparison across websites[a]		η_p^2
	Mean (SD)	n (%)	Mean (SD)	n (%)	*F* test (*df*)	*P* value	
Knowledge[b]	5.60 (2.24)	174 (52)	8.28 (1.76)	158 (48)	173.821 (1,329)	<.001	0.346
Attitudes[c]	3.89 (1.28)	173 (53)	3.24 (1.31)	156 (47)	15.779 (1,326)	<.001	0.046
Intentions[c]	2.58 (1.65)	172 (52)	3.34 (1.87)	161 (48)	13.264 (1,330)	<.001	0.039

[a]Models adjusted for baseline knowledge.

[b]Possible score range 0 to 10 (10=high knowledge).

[c]Possible score range 1 to 7 (7=positive attitudes or intentions).

Table 4. Informed choice categories.

Informed choice	Knowledge	Attitude	Intentions	Whole sample (N=314), n (%)	Control website (n=168), n (%)	PoP[a] (n=146), n (%)
No	Low	Positive	Negative	28 (8.9)	25 (14.9)	3 (2.1)
No	Low	Negative	Positive	17 (5.4)	14 (8.3)	3 (2.1)
No	Low	Positive	Positive	29 (9.2)	20 (11.9)	9 (6.2)
No	Low	Negative	Negative	86 (27.4)	69 (41.1)	17 (11.6)
No	High	Positive	Negative	50 (15.9)	11 (6.5)	39 (26.7)
No	High	Negative	Positive	14 (4.5)	5 (3.0)	9 (6.2)
Yes	High	Positive	Positive	52 (16.6)	9 (5.4)	43 (29.5)
Yes	High	Negative	Negative	38 (12.1)	15 (8.9)	23 (15.8)

[a]PoP: The Power of Placebos.

Table 5. Postintervention treatment beliefs by group.

Treatment belief	Control website		PoP[a]		Comparison across websites[b]		η_p^2
	Mean (SD)	n (%)	Mean (SD)	n (%)	*F* test (*df*)	*P* value	
Concerns[c]	2.40 (0.82)	173 (51)	2.29 (0.84)	163 (49)	0.810 (1,333)	0.37	0.002
Individual fit[c]	2.19 (0.91)	172 (52)	2.76 (1.03)	159 (48)	23.728 (1,328)	<.001	0.067
Expectancy[c]	2.59 (1.00)	172 (52)	3.40 (0.88)	161 (48)	58.657 (1,330)	<.001	0.151
Credibility[c]	2.43 (0.90)	173 (52)	3.06 (0.91)	162 (48)	36.529 (1,332)	<.001	0.099

[a]PoP: The Power of Placebos.

[b]Models adjusted for baseline knowledge.

[c]Possible score range 1 to 5 (5=positive beliefs or fewer concerns about placebo).

Discussion

Principal Findings

This study tested the effects of a new person-based website about placebo effects, PoP, on adults with recent back pain, by comparing it with a control website based on existing written UK patient information leaflets. Participants who viewed PoP had greater increases in knowledge about placebos and were 3 times more likely to make an informed choice about placebos compared with participants who viewed the control website. On average, participants who viewed PoP answered 2 more knowledge quiz items correctly (out of a total of 10 items). Compared with the control website, PoP also led participants to perceive placebos as more credible, more effective, and more suitable for them personally and to be more willing to try a placebo in the future.

Other studies have also successfully modified people's beliefs and/or knowledge about placebos. For example, 1 study [22] compared a brief educational intervention comprising 5 slides presenting information about placebo effects and their mechanisms with a control intervention presenting information about the epidemiology, costs, and risk factors for musculoskeletal pain. Participants with chronic musculoskeletal pain were randomly assigned to 1 of the 2 interventions. Those who viewed the placebo educational intervention subsequently reported feeling more knowledgeable about placebo analgesia,

seeing placebos as more active, more effective, and more acceptable in a range of hypothetical clinical scenarios [22]. Another study [23] compared 2 patient information leaflets about placebo-controlled randomized trials, one based on standard information and the other supplemented with additional information about placebos, their effects, and mechanisms of action. In an online randomized experiment, people with chronic illness who viewed the supplemented leaflet subsequently reported significantly higher expectations and perceptions of the credibility of placebo treatment for pain compared with those who viewed the standard information leaflet [23]. PoP is more comprehensive than educational and/or informational resources on placebo effects reported previously, and it was developed using a systematic approach described in detail elsewhere [28]. Similar to the studies by Kisaalita et al [22] and Bishop et al [23], this study has shown that information about placebos and placebo effects can lead patients to have more positive beliefs about placebos. This study has also shown effects on beliefs about placebos in a large sample of patients with a particular pain condition. Uniquely, PoP had a significant effect on objectively measured knowledge and informed choice, a close analogy to informed consent in both clinical and research settings. PoP also increased willingness to try a placebo, suggesting it might help facilitate recruitment to placebo-controlled randomized controlled trials (RCTs). Future work should test PoP in the context of a placebo-controlled trial, to ascertain whether it can indeed encourage more people to volunteer.

PoP could be readily adapted for use in other conditions, which, like pain, have well-documented placebo effects (such as irritable bowel syndrome [53] or depression [54]), and/or specific placebo-controlled RCTs. PoP could be adapted to improve patients' understanding of open-label placebo interventions [35,55] and/or to educate patients about placebo effects and mind-body interactions more broadly. It would be interesting to examine its impact on willingness to use such interventions. It would also be interesting to explore its use in placebo-controlled surgery trials where there are large placebo effects [56]. However, it would not be appropriate to use PoP as part of participant information for trials in conditions in which placebo effects are poorly understood or rarely seen, making it important to consult up-to-date reviews of placebo effects before finalizing placebo information for any specific trial. Furthermore, before it is used in an RCT, it is important to test for any additional effects of PoP on trial variables, specifically the size of the placebo effect in both placebo and verum arms, the success of blinding, and patients' reactions to debriefing. In this study, the person-based website led to increased expectations of effectiveness of a placebo. Positive outcome expectations are a key mechanism underpinning placebo effects in pain and other conditions [31]. Therefore, if PoP were to increase patients' expectations of effectiveness in a placebo-controlled RCT, this could also lead to larger placebo effects and thus have implications for power calculations and the efficacy of the target treatment.

The use of PoP in RCTs could also have implications for successful blinding of patients to treatment allocation. Qualitative studies suggest that patients typically believe if they experience no side effects in a trial, then this means they are receiving the placebo, whereas if they do experience side effects, then this means they are receiving the verum intervention [12,57]. This logic is appealing but incorrect as patients in placebo groups regularly report benefits and adverse effects, particularly in pain trials [5,58]. As PoP increases patients' knowledge about the positive and adverse effects of placebos, this could reduce patients' confidence in a link between effects and treatment allocation, thus helping to maintain uncertainty and, therefore, blinding. Believing that placebos can have no effects might also lead patients to be surprised and/or distressed on being debriefed at the end of a trial and finding out they were in the placebo group [10,14,15,57]. Future research could use PoP to test whether providing more comprehensive information about placebos at the start of a trial can improve patients' and investigators' experiences of debriefing at the end of a trial.

Strengths and Limitations

Strengths of this study include the systematic approach used to develop the person-based website, the choice of a Web-based modality and the use of a control intervention. PoP had previously been developed using person-based, theory-based, and evidence-based intervention designs [28,42]. This ensured it was engaging, persuasive, and based on current scientific evidence about the size and mechanisms of action of placebo effects. Compared with traditional paper-based patient information leaflets, creating a website enabled us to provide more detailed information in an accessible and engaging manner using a range of formats including text, audio, and film. Other studies of person-based digital interventions typically compare them with usual care (ie, nondigital) [59,60]. This study contributes a demonstration of the value of the person-based approach to intervention development as compared with a website based on standard written patient information.

Limitations stem from the choice of control, the sample, and the measure of informed choice. The use of a control website based on existing printed materials enabled a pragmatic evaluation of the potential impact of PoP but did not allow an evaluation of which components of the website were most important; for example, both websites included (different) text and images, but PoP also included audio clips and short films and was longer than the control website. It is, therefore, possible that the effects of PoP were because of these differences in sheer volume of information.

Online health information is accessible to a large majority—but not all—of the population: in 2015, 86% of UK households had internet access and 78% of adults accessed the internet daily or almost daily [61]. The generalizability of the study is somewhat limited as there was a very low uptake from the initial mail out via GP surgeries (leading to possible selection bias) and only people with self-reported recent back pain were included. It is not possible to compare the participants with all patients who were invited to take part as no information could be ascertained about the latter group. As we invited patients who had consulted their GP with back pain as much as 3 years ago, it is possible that many of the nonresponders were no longer seeking treatment for back pain, and therefore, this study was not relevant to them. The inclusion criteria do not directly map onto

current definitions of acute or chronic back pain but were chosen to capture a group of people for whom PoP might be of interest. The website should have similar effects on other groups of people with painful conditions, but this could be tested in future studies. The final sample size fell short of the a priori sample size calculation by 1 participant, but as almost all the results were statistically significant, this did not seem to have an impact on the findings.

The outcome measures were previously validated and included an assessment of objective knowledge and an assessment of informed choice, which is particularly relevant when considering the potential use of the website to inform volunteers for RCTs. However, the conceptual strength of measuring informed choice must be balanced by an acknowledgment of the statistical limitations of the loss of data associated with this particular outcome measure, as it was derived (following published guidelines) by dichotomizing 3 continuous variables. Finally, it must be noted that the participants in this study were asked about intentions and willingness to take part in a placebo-controlled trial without actually being invited to take part in a specific trial. This is a limitation for at least two reasons. First, intentions and willingness to do something do not always translate into actual behavior [62]. Second, placebo-controlled trial participants receive information about the trial treatment and procedures as well as about the placebo, and they have to decide whether they would be willing to receive either, not just one, of these interventions. Future studies should, therefore, examine the use of enhanced information resources such as the PoP website in the context of an actual placebo-controlled trial.

Conclusions

In conclusion, the person-based website, PoP, increased knowledge, led more participants to make an informed choice, and enhanced positive beliefs about placebos in a sample of adults with recent back pain recruited from primary care. It could be used to increase levels of understanding about placebo effects among the general public. In future, PoP could be adapted and used to support informed consent and recruitment to placebo-controlled clinical trials, but first, its effects on recruitment and trial outcomes should be investigated in the context of a placebo-controlled trial. This study provides initial evidence suggesting that the person-based method of developing Web-based interventions—combining extensive qualitative research with evidence-based and theory-based methods [42]—could be used to improve informed consent materials in clinical research. Further work is needed to confirm its utility in the development of materials to support the ethical conduct of clinical research.

Acknowledgments

This project was funded by a grant from Arthritis Research UK, Chesterfield, UK (grant reference 20113). Additional funding was provided by Solent NHS Trust (Research Capacity Funding). The funders and study sponsors (University of Southampton) had no role in the design of the study; collection, analysis, and interpretation of data; and writing the manuscript. The authors would like to thank all those who participated in the trial and the primary care research network and GP surgeries who facilitated recruitment. Christine Bagg (patient advisor) contributed to study design and helped review the website. Dr Jin Zhang and the core LifeGuide team provided technical support for programming the Web page.

Conflicts of Interest

The authors both developed and evaluated PoP and the control website. They have no other conflicts of interest.

Editorial notice: This randomized study was not prospectively registered, justified by the authors as “it does not measure a health outcome”. While educational outcomes related to health interventions/topics can be considered health outcomes, 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 their primary outcomes or effectiveness, as the lack of registration means that authors could change their outcome measures retrospectively. Best practice is to register trials and/or publish protocols prospectively.

Multimedia Appendix 1

Description of the factorial trial design.

[PDF File (Adobe PDF File), 11KB - jmir_v21i1e9955_app1.pdf]

Multimedia Appendix 2

Screenshots showing additional pages from PoP.

[PDF File (Adobe PDF File), 751KB - jmir_v21i1e9955_app2.pdf]

Multimedia Appendix 3

A completed TIDieR checklist with full information about the description of PoP.

[PDF File (Adobe PDF File), 67KB - jmir_v21i1e9955_app3.pdf]

Multimedia Appendix 4

Screenshots of the entire control website.

[PDF File (Adobe PDF File), 820KB - jmir_v21i1e9955_app4.pdf]

Multimedia Appendix 5

Information sheet and consent form for study participants.

[PDF File (Adobe PDF File), 108KB - jmir_v21i1e9955_app5.pdf]

References

1. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. J Am Med Assoc 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
2. Bishop FL, Adams AE, Kaptchuk TJ, Lewith GT. Informed consent and placebo effects: a content analysis of information leaflets to identify what clinical trial participants are told about placebos. PLoS One 2012 Jun;7(6):e39661 [FREE Full text] [doi: 10.1371/journal.pone.0039661] [Medline: 22761859]
3. Finniss DG, Kaptchuk TJ, Miller F, Benedetti F. Biological, clinical, and ethical advances of placebo effects. Lancet 2010 Feb 20;375(9715):686-695 [FREE Full text] [doi: 10.1016/S0140-6736(09)61706-2] [Medline: 20171404]
4. Benedetti F, Amanzio M. The placebo response: how words and rituals change the patient's brain. Patient Educ Couns 2011 Sep;84(3):413-419. [doi: 10.1016/j.pec.2011.04.034] [Medline: 21621366]
5. Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. Cochrane Database Syst Rev 2010(1):CD003974. [doi: 10.1002/14651858.CD003974.pub3] [Medline: 20091554]
6. Forsberg JT, Martinussen M, Flaten MA. The placebo analgesic effect in healthy individuals and patients: a meta-analysis. Psychosom Med 2017 May;79(4):388-394. [doi: 10.1097/PSY.0000000000000432] [Medline: 27922566]
7. Howick J, Friedemann C, Tsakok M, Watson R, Tsakok T, Thomas J, et al. Are treatments more effective than placebos? A systematic review and meta-analysis. PLoS One 2013;8(5):e62599 [FREE Full text] [doi: 10.1371/journal.pone.0062599] [Medline: 23690944]
8. Blease CR, Bishop FL, Kaptchuk TJ. Informed consent and clinical trials: where is the placebo effect? Br Med J 2017 Feb 03;356:j463. [doi: 10.1136/bmj.j463] [Medline: 28159769]
9. Criscione LG, Sugarman J, Sanders L, Pisetsky DS, St Clair EW. Informed consent in a clinical trial of a novel treatment for rheumatoid arthritis. Arthritis Rheum 2003 Jun 15;49(3):361-367 [FREE Full text] [doi: 10.1002/art.11057] [Medline: 12794792]
10. Bishop FL, Jacobson EE, Shaw JR, Kaptchuk TJ. Scientific tools, fake treatments, or triggers for psychological healing: how clinical trial participants conceptualise placebos. Soc Sci Med 2012 Mar;74(5):767-774 [FREE Full text] [doi: 10.1016/j.socscimed.2011.11.020] [Medline: 22285289]
11. Kaptchuk TJ, Shaw J, Kerr CE, Conboy LA, Kelley JM, Csordas TJ, et al. "Maybe I made up the whole thing": placebos and patients' experiences in a randomized controlled trial. Cult Med Psychiatry 2009 Sep;33(3):382-411 [FREE Full text] [doi: 10.1007/s11013-009-9141-7] [Medline: 19597976]
12. White P, Bishop FL, Prescott P, Scott C, Little P, Lewith G. Practice, practitioner, or placebo? A multifactorial, mixed-methods randomized controlled trial of acupuncture. Pain 2012 Feb;153(2):455-462. [doi: 10.1016/j.pain.2011.11.007] [Medline: 22169359]
13. Welton AJ, Vickers MR, Cooper JA, Meade TW, Marteau TM. Is recruitment more difficult with a placebo arm in randomised controlled trials? A quasirandomised, interview based study. Br Med J 1999 Apr 24;318(7191):1114-1117 [FREE Full text] [Medline: 10213724]
14. Bishop FL, Jacobson EE, Shaw J, Kaptchuk TJ. Participants' experiences of being debriefed to placebo allocation in a clinical trial. Qual Health Res 2012 Aug;22(8):1138-1149 [FREE Full text] [doi: 10.1177/1049732312448544] [Medline: 22673094]
15. Di Blasi Z, Crawford F, Bradley C, Kleijnen J. Reactions to treatment debriefing among the participants of a placebo controlled trial. BMC Health Serv Res 2005 Apr 22;5(1):30 [FREE Full text] [doi: 10.1186/1472-6963-5-30] [Medline: 15847683]
16. Chen G, Johnson MH. Patients' attitudes to the use of placebos: results from a New Zealand survey. N Z Med J 2009 Jun 05;122(1296):35-46. [Medline: 19652679]
17. Berthelot JM, Maugars Y, Abgrall M, Prost A. Interindividual variations in beliefs about the placebo effect: a study in 300 rheumatology inpatients and 100 nurses. Joint Bone Spine 2001 Feb;68(1):65-70. [Medline: 11235784]

18. Hughes J, Greville-Harris M, Graham CA, Lewith G, White P, Bishop FL. What trial participants need to be told about placebo effects to give informed consent: a survey to establish existing knowledge among patients with back pain. J Med Ethics 2017 Dec;43(12):867-870 [FREE Full text] [doi: 10.1136/medethics-2016-103964] [Medline: 28663259]
19. Keränen T, Halkoaho A, Itkonen E, Pietilä A. Placebo-controlled clinical trials: how trial documents justify the use of randomisation and placebo. BMC Med Ethics 2015 Jan 11;16:2 [FREE Full text] [doi: 10.1186/1472-6939-16-2] [Medline: 25578433]
20. Hernández A, Baños J, Llop C, Farré M. The definition of placebo in the informed consent forms of clinical trials. PLoS One 2014;9(11):e113654 [FREE Full text] [doi: 10.1371/journal.pone.0113654] [Medline: 25423149]
21. Cheon S, Park H, Chae Y, Lee H. Does different information disclosure on placebo control affect blinding and trial outcomes? A case study of participant information leaflets of randomized placebo-controlled trials of acupuncture. BMC Med Res Methodol 2018 Jan 18;18(1):13 [FREE Full text] [doi: 10.1186/s12874-018-0474-1] [Medline: 29347917]
22. Kisaalita NR, Hurley RW, Staud R, Robinson ME. Placebo use in pain management: a mechanism-based educational intervention enhances placebo treatment acceptability. J Pain 2016 Feb;17(2):257-269 [FREE Full text] [doi: 10.1016/j.jpain.2015.10.017] [Medline: 26604097]
23. Bishop FL, McGinn L, Graham C, Biggs H, Denegri S, Lewith GT. Describing placebos in patient information leaflets: effects on patients' beliefs. Eur J Pers Cent Healthc 2016;4(4):609-617. [doi: 10.5750/ejpch.v4i4.1145]
24. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki CE, Prokosch H, Wynn R. Informed citizen and empowered citizen in health: results from an European survey. BMC Fam Pract 2011;12:20 [FREE Full text] [doi: 10.1186/1471-2296-12-20] [Medline: 21496309]
25. Atkinson NL, Saperstein SL, Pleis J. Using the internet for health-related activities: findings from a national probability sample. J Med Internet Res 2009;11(1):e4 [FREE Full text] [doi: 10.2196/jmir.1035] [Medline: 19275980]
26. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
27. Cook DA, Dupras DM. A practical guide to developing effective web-based learning. J Gen Intern Med 2004 Jun;19(6):698-707 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30029.x] [Medline: 15209610]
28. Greville-Harris M, Bostock J, Din A, Graham CA, Lewith G, Liossi C, et al. Informing patients about placebo effects: using evidence, theory, and qualitative methods to develop a new website. JMIR Res Protoc 2016 Jun 10;5(2):e106 [FREE Full text] [doi: 10.2196/resprot.5627] [Medline: 27288271]
29. Mistiaen P, van Osch M, van Vliet L, Howick J, Bishop FL, DiBlasi Z, et al. The effect of patient-practitioner communication on pain: a systematic review. Eur J Pain 2015 Oct 22;20(5):675-688. [doi: 10.1002/ejp.797] [Medline: 26492629]
30. Zhang W, Robertson J, Jones AC, Dieppe PA, Doherty M. The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. Ann Rheum Dis 2008 Dec;67(12):1716-1723. [doi: 10.1136/ard.2008.092015] [Medline: 18541604]
31. Benedetti F. Mechanisms of placebo and placebo-related effects across diseases and treatments. Annu Rev Pharmacol Toxicol 2008;48:33-60. [doi: 10.1146/annurev.pharmtox.48.113006.094711] [Medline: 17666008]
32. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. Lancet 2018 Mar 20;391(10137):2356-2367. [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
33. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Dec 08;388(10053):1545-1602 [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
34. Foster NE, Anema JR, Cherkin D, Chou R, Cohen SP, Gross DP, Lancet Low Back Pain Series Working Group. Prevention and treatment of low back pain: evidence, challenges, and promising directions. Lancet 2018 Jun 09;391(10137):2368-2383. [doi: 10.1016/S0140-6736(18)30489-6] [Medline: 29573872]
35. Carvalho C, Caetano JM, Cunha L, Rebouta P, Kaptchuk TJ, Kirsch I. Open-label placebo treatment in chronic low back pain: a randomized controlled trial. Pain 2016 Dec;157(12):2766-2772 [FREE Full text] [doi: 10.1097/j.pain.0000000000000700] [Medline: 27755279]
36. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. Health Expect 2014 Oct;17(5):637-650. [doi: 10.1111/j.1369-7625.2012.00795.x] [Medline: 22809132]
37. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [Medline: 17695343]
38. Williams S, Yardley L, Wills GB. A qualitative case study of LifeGuide: users' experiences of software for developing Internet-based behaviour change interventions. Health Informatics J 2013 Mar;19(1):61-75. [doi: 10.1177/1460458212458915] [Medline: 23486826]
39. Coulter A, Ellins J. The Health Foundation. 2006. Patient-focused interventions. A review of the evidence URL: https://www.health.org.uk/publications/patient-focused-interventions [accessed 2019-01-03] [WebCite Cache ID 759qNtx2x]

40. Michie S, Abraham C. Interventions to change health behaviors: evidence-based or evidence-inspired? Psychol Health 2004;19(1):29-49. [doi: 10.1080/0887044031000141199]
41. Michie S, Johnston M. Theories and techniques of behaviour change: developing a cumulative science of behaviour change. Health Psychol Rev 2012 Mar;6(1):1-6. [doi: 10.1080/17437199.2012.654964]
42. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
43. Price DD, Finniss DG, Benedetti F. A comprehensive review of the placebo effect: recent advances and current thought. Annu Rev Psychol 2008;59:565-590. [doi: 10.1146/annurev.psych.59.113006.095941] [Medline: 17550344]
44. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
45. Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, et al. Report of the NIH Task Force on research standards for chronic low back pain. J Pain 2014 Jun;15(6):569-585 [FREE Full text] [doi: 10.1016/j.jpain.2014.03.005] [Medline: 24787228]
46. Marteau TM, Dormandy E, Michie S. A measure of informed choice. Health Expect 2001 Jun;4(2):99-108. [Medline: 11359540]
47. Dormandy E, Hooper R, Michie S, Marteau TM. Informed choice to undergo prenatal screening: a comparison of two hospitals conducting testing either as part of a routine visit or requiring a separate visit. J Med Screen 2002;9(3):109-114. [Medline: 12370321]
48. Michie S, Dormandy E, Marteau TM. The multi-dimensional measure of informed choice: a validation study. Patient Educ Couns 2002 Sep;48(1):87-91. [Medline: 12220754]
49. Francis J, Eccles M, Johnston M, Walker A, Grimshaw J, Foy R, et al. City Research Online. 2004. Constructing questionnaires based on the theory of planned behaviour. A manual for health services researchers URL: http://openaccess.city.ac.uk/1735/ [accessed 2018-11-20] [WebCite Cache ID 744GrqB22]
50. Dima A, Lewith GT, Little P, Moss-Morris R, Foster NE, Hankins M, et al. Patients' treatment beliefs in low back pain: development and validation of a questionnaire in primary care. Pain 2015 Aug;156(8):1489-1500. [doi: 10.1097/j.pain.0000000000000193] [Medline: 25906346]
51. Tabachnick BG, Fidell LS. Using Multivariate Statistics. Boston: Allyn & Bacon; 2001.
52. Van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. J Clin Epidemiol 2006 Sep;59(9):920-925. [doi: 10.1016/j.jclinepi.2006.02.007] [Medline: 16895814]
53. Patel SM, Stason WB, Legedza A, Ock SM, Kaptchuk TJ, Conboy L, et al. The placebo effect in irritable bowel syndrome trials: a meta-analysis. Neurogastroenterol Motil 2005 Jun;17(3):332-340. [doi: 10.1111/j.1365-2982.2005.00650.x] [Medline: 15916620]
54. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med 2008 Feb;5(2):e45 [FREE Full text] [doi: 10.1371/journal.pmed.0050045] [Medline: 18303940]
55. Kaptchuk TJ, Friedlander E, Kelley JM, Sanchez MN, Kokkotou E, Singer JP, et al. Placebos without deception: a randomized controlled trial in irritable bowel syndrome. PLoS One 2010;5(12):e15591 [FREE Full text] [doi: 10.1371/journal.pone.0015591] [Medline: 21203519]
56. Jonas WB, Crawford C, Colloca L, Kaptchuk TJ, Moseley B, Miller FG, et al. To what extent are surgery and invasive procedures effective beyond a placebo response? A systematic review with meta-analysis of randomised, sham controlled trials. BMJ Open 2015 Dec 11;5(12):e009655 [FREE Full text] [doi: 10.1136/bmjopen-2015-009655] [Medline: 26656986]
57. Kwakye IN, Garner M, Baldwin DS, Bamford S, Pinkney V, Bishop FL. Altruism, personal benefit, and anxieties: a phenomenological study of healthy volunteers' experiences in a placebo-controlled trial of duloxetine. Hum Psychopharmacol 2016 Dec;31(4):332-340 [FREE Full text] [doi: 10.1002/hup.2543] [Medline: 27378326]
58. Papadopoulos D, Mitsikostas DD. A meta-analytic approach to estimating nocebo effects in neuropathic pain trials. J Neurol 2012 Mar;259(3):436-447. [doi: 10.1007/s00415-011-6197-4] [Medline: 21811804]
59. Geraghty AW, Kirby S, Essery R, Little P, Bronstein A, Turner D, et al. Internet-based vestibular rehabilitation for adults aged 50 years and over: a protocol for a randomised controlled trial. BMJ Open 2014 Jul 22;4(7):e005871 [FREE Full text] [doi: 10.1136/bmjopen-2014-005871] [Medline: 25052178]
60. Little P, Stuart B, Francis N, Douglas E, Tonkin-Crine S, Anthierens S, et al. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: a multinational, cluster, randomised, factorial, controlled trial. Lancet 2013 Oct 5;382(9899):1175-1182 [FREE Full text] [doi: 10.1016/S0140-6736(13)60994-0] [Medline: 23915885]
61. Office for National Statistics. Office for National Statistics. 2015 Aug 06. Statistical Bulletin: Internet Access - Households and Individuals URL: http://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2015-08-06 [accessed 2018-11-19] [WebCite Cache ID 6g1la5vII]
62. Sheeran P, Webb TL. The intention-behavior gap. Soc Personal Psychol Compass 2016 Sep 05;10(9):503-518. [doi: 10.1111/spc3.12265]

Abbreviations

ANOVA: analysis of variance
GP: general practitioner
MCAR: missing completely at random
PoP: The Power of Placebos
RCT: randomized controlled trial
TIDieR: template for intervention description and replication

Edited by G Eysenbach; submitted 06.02.18; peer-reviewed by M Alshehri, S Atlas; comments to author 28.07.18; revised version received 28.08.18; accepted 04.10.18; published 17.01.19.

Please cite as:
Bishop FL, Greville-Harris M, Bostock J, Din A, Graham CA, Lewith G, Liossi C, O'Riordan T, White P, Yardley L
Informing Adults With Back Pain About Placebo Effects: Randomized Controlled Evaluation of a New Website With Potential to Improve Informed Consent in Clinical Research
J Med Internet Res 2019;21(1):e9955
URL: https://www.jmir.org/2019/1/e9955/
doi:10.2196/jmir.9955
PMID:30664462

Original Paper

Improving Electronic Health Record Note Comprehension With NoteAid: Randomized Trial of Electronic Health Record Note Comprehension Interventions With Crowdsourced Workers

John P Lalor[1], MS; Beverly Woolf[1], PhD; Hong Yu[1,2,3,4], PhD, FACMI

[1]College of Information and Computer Sciences, University of Massachusetts, Amherst, MA, United States

[2]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States

[3]Department of Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[4]Bedford Veterans Affairs Medical Center, Center for Healthcare Organization and Implementation Research, Bedford, MA, United States

Corresponding Author:
Hong Yu, PhD, FACMI
Department of Computer Science
University of Massachusetts Lowell
1 University Avenue
Lowell, MA,
United States
Phone: 1 508 612 7292
Email: hong.yu@umassmed.edu

Abstract

Background: Patient portals are becoming more common, and with them, the ability of patients to access their personal electronic health records (EHRs). EHRs, in particular the free-text EHR notes, often contain medical jargon and terms that are difficult for laypersons to understand. There are many Web-based resources for learning more about particular diseases or conditions, including systems that directly link to lay definitions or educational materials for medical concepts.

Objective: Our goal is to determine whether use of one such tool, NoteAid, leads to higher EHR note comprehension ability. We use a new EHR note comprehension assessment tool instead of patient self-reported scores.

Methods: In this work, we compare a passive, self-service educational resource (MedlinePlus) with an active resource (NoteAid) where definitions are provided to the user for medical concepts that the system identifies. We use Amazon Mechanical Turk (AMT) to recruit individuals to complete ComprehENotes, a new test of EHR note comprehension.

Results: Mean scores for individuals with access to NoteAid are significantly higher than the mean baseline scores, both for raw scores (P=.008) and estimated ability (P=.02).

Conclusions: In our experiments, we show that the active intervention leads to significantly higher scores on the comprehension test as compared with a baseline group with no resources provided. In contrast, there is no significant difference between the group that was provided with the passive intervention and the baseline group. Finally, we analyze the demographics of the individuals who participated in our AMT task and show differences between groups that align with the current understanding of health literacy between populations. This is the first work to show improvements in comprehension using tools such as NoteAid as measured by an EHR note comprehension assessment tool as opposed to patient self-reported scores.

(J Med Internet Res 2019;21(1):e10793) doi:10.2196/10793

KEYWORDS

health literacy; crowdsourcing; natural language processing; information storage and retrieval; psychometrics; MedlinePlus

Introduction

Background and Significance

In recent years, many hospitals have adopted patient portals to make medical records available to patients. In particular, patient portals allow patients to access their electronic health records (EHRs). In a survey of studies related to patient access to their medical records, generally, patients who chose to see their records were satisfied with their contents [1-4] and felt greater autonomy about their care [1,5,6]. Granting patients access to

their records also does not increase the workload of medical staff members [1,7-9]. Generally, patient access to EHRs can lead to positive health outcomes and greater understanding of their conditions [1,10,11]. However, EHRs and the progress notes that are included often contain complex medical jargon that is difficult for patients to comprehend. When given access to their notes, patients have questions about the meaning of medical terms and other concepts included in the notes [9,12]. Tools such as OpenNotes have promoted the inclusion of patient visit notes in patient portals, but simply including the notes may not be beneficial for patients if they have questions regarding the meaning of terms in the notes. Tools and resources that can define terms and provide lay definitions for medical concepts are needed as part of the move to make EHR notes available to patients so that they can understand the contents of their notes and their medical record.

Self-service educational materials are widely available, especially on the Web. There is a wealth of information related to medicine and health care on the internet, ranging from well-maintained ontologies with curated educational materials to Web-based discussion communities of patients that suffer from the same disease. With this information, patients with certain symptoms can find information about their condition on the internet. But is the wealth of information useful? That is, does simply having access to health information lead to better understanding? In this work, we test the usefulness of both passive and active interventions for assisting patients with understanding medical concepts. The passive system, MedlinePlus (MLP) [13], is a Web-based repository maintained by the US National Library of Medicine that includes information and definitions for clinical concepts, diseases, and other terms related to health care. MLP has been used in the past to promote patient education and provide patients with definitions and educational material to improve health literacy [14-17]. MLP is a large repository of high-quality health care information, but the user must search for the information that he or she is looking for. MLP does not automatically surface information for users.

NoteAid [18,19] is a freely available Web-based system developed by our team that automatically identifies medical concepts and displays their definitions to users. NoteAid has previously been shown to improve patients' understanding of notes as measured by self-reporting [18,19].

In this work, our goal is to determine if access to NoteAid or MLP is associated with higher levels of EHR note comprehension. Do these interventions of educational materials improve a patient's ability to comprehend his or her EHR note? In this work, we use the Amazon Mechanical Turk (AMT) microtask crowdsourcing platform to give AMT workers (Turkers) the ComprehENotes EHR note comprehension test [20], a set of questions designed to test EHR note comprehension. AMT is an increasingly popular tool for gathering research data [21-23] and recruiting participants for experiments, both in open-domain tasks [24,25] and medical-specific research [26-29]. Certain Turkers were not given 1 of the external resources, whereas others were provided with either MLP or NoteAid. Our results show that using NoteAid leads to significantly higher scores on the EHR comprehension test compared with the baseline population that was given no external resource. However, we found no significant difference between the Turkers with no resource and the Turkers who used MLP. Turkers were also asked to take the short Test of Functional Health Literacy in Adults (S-TOFHLA) to assess functional health literacy. All the Turkers scored *adequate health literacy*, the highest level for S-TOFHLA. This is the first work to quantitatively analyze the impact of tools such as NoteAid using a test of EHR note comprehension as opposed to self-reported scores.

In this work, we show that NoteAid has a significant impact on EHR note comprehension as measured by a test specific to that task. In addition, simply giving a patient access to sites such as MLP does not lead to significant improvements in test scores over a baseline group that had no external resources available to them. Finally, we analyze the demographics of the Turkers who completed our tasks. A regression model to predict test scores showed differences between demographic groups that align with the current knowledge regarding health literacy. For example, individuals that reported education of less than high school scored lower than average, whereas individuals that identified as white scored higher than average.

Related Work

Health literacy is an important issue for patients. Low health literacy is a widespread problem, with only 12% of adults estimated to be proficient in health literacy [30]. The Institute of Medicine defines health literacy as "the degree to which individuals have the capacity to make appropriate decisions regarding their health" [31]. Patients with low health literacy often have difficulty with understanding instructions for medications from their doctors and have trouble navigating systems for making appointments, filling prescriptions, and fulfilling other health-related tasks [32,33]. In addition, having low health literacy has been linked to negative health outcomes in areas such as heart disease and fear of cancer progression [34,35].

It is important to be able to test a patient's health literacy to identify those patients with low health literacy. Doctors can then provide these patients with educational materials to improve their understanding of medical terms and concepts. Testing health literacy is especially important with the proliferation of Web-based patient portals, where patients can access their EHRs and EHR notes directly. Giving a patient access to their EHRs and EHR notes without confirming that the patient can understand the content of the notes may lead to confusion and frustration with their health care experience.

There are a number of tests for health literacy, including the Test of Functional Health Literacy in Adults (TOFHLA) and the Newest Vital Sign (NVS) [36-38]. TOFHLA and its shortened form (S-TOFHLA) test comprehension and numeracy by providing scenarios to patients and constructing fill-in-the-blank questions by removing key terms from the scenario passages. NVS is a short test where patients are required to answer questions related to a nutrition label, to test whether the patient can navigate the label. These tests work well as screening instruments to identify patients who may have

low health literacy, but they are broad tests and do not specifically test EHR note comprehension.

Although these and other tests are available, the only test that specifically targets a patient's ability to comprehend their EHR notes is the ComprehENotes test [20]. The ComprehENotes test questions were developed using key concepts extracted from deidentified EHR notes. Questions were written by physicians and medical researchers using Sentence Verification Technique and validated using Item Response Theory (IRT) [39,40]. The test set is the first of its kind that specifically tests a patient's ability to comprehend the type of content that is included in EHR notes.

Methods

Overview

In this work, we recruited Turkers on the AMT platform and asked them to complete the ComprehENotes EHR note comprehension test. Turkers were split into 3 groups and were allowed to use 1 external resource when completing the test (or no resource in the case of the baseline group). Test results were collected and analyzed using IRT to estimate EHR note comprehension ability for each of the individuals, and group results were analyzed to determine if either of the external resources had a significant effect on test scores. Figure 1 illustrates our methodology at a high level. Details for each of the steps are described below.

Data Collection

To assess EHR note comprehension, we used the ComprehENotes question set [20]. The dataset consists of 55 questions to measure EHR note comprehension. A bank of 154 questions was developed by groups of physicians and medical researchers from deidentified patient notes and then filtered down to a final test set using IRT. A total of 83 of the 154 questions were provided to AMT Turkers, who provided responses. These responses were used to fit an IRT model that estimated the questions' ability to test EHR note comprehension. Of the questions in the original question bank, 55 were retained as a test of note comprehension [20].

The questions in the ComprehENotes test set include questions from patient EHR notes associated with 6 diseases: heart failure, hypertension, diabetes, chronic obstructive pulmonary disease (COPD), liver failure, and cancer. The questions are all general enough that they assess a key concept associated with 1 of the 6 diseases without being so specific to a single patient that they are not useful to others [20]. Therefore, the test can be used to assess a patient's general EHR note comprehension ability and allows for comparisons between patients with respect to comprehension ability.

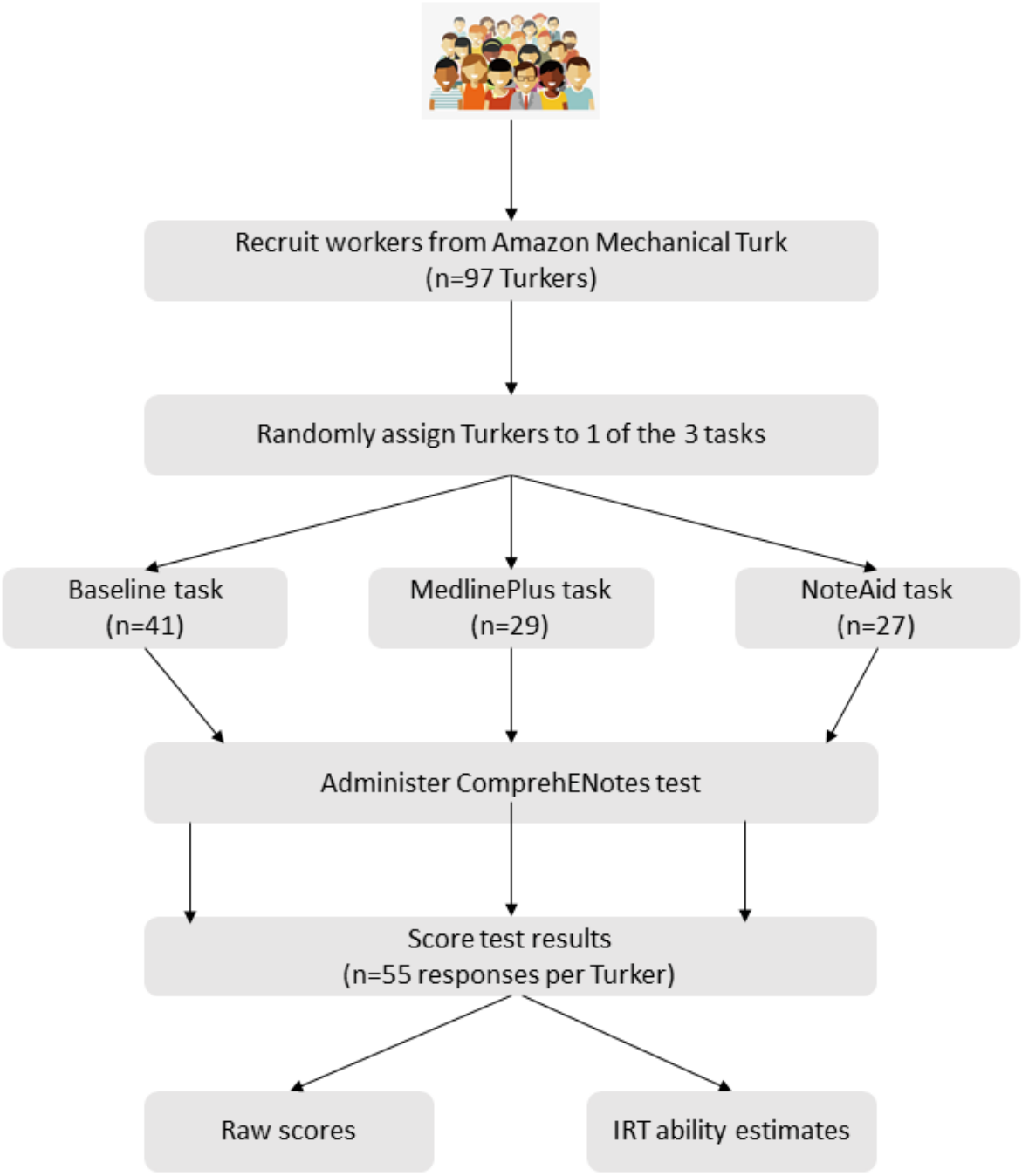
The ComprehENotes test set is most informative for individuals with low health literacy. That is, the SE of the ability estimation is lowest at low levels of ability (eg, −2 to −0.5). In addition, most of the ComprehENotes questions have low difficulty parameters. The difficulty parameters range from −2.2 to 0.7. That is, the questions are of a difficulty that individuals with lower than average ability have a 50% chance of answering correctly. For example, if a question has a difficulty parameter of −1.0, then an individual with estimated ability of −1.0 has a 50% chance of answering the question correctly. Ability estimates are normally distributed, so an individual with estimated ability of −1.0 is 1 SD below the average individual. Textbox 1 shows two example questions taken from the ComprehENotes test. Individuals are shown a snippet of text from a deidentified EHR note and asked to select the answer that has the same meaning as the italicized portion of the text.

We set up 3 AMT tasks for Turkers to complete. Turkers were presented with the ComprehENotes question set, 1 question at a time, and were asked to provide the correct answer.

For 1 task (Baseline), the Turkers were instructed to not use any external resources when answering the questions. For the first treatment task (Treatment-MLP), Turkers were given a link to MLP and were told that they could use the site as a reference when completing the task. Turkers were encouraged to use the MLP page search functionality to search for definitions to unknown terms or concepts that appeared in the task. For the second treatment task (Treatment-NoteAid [Treatment-NA]), the Turkers were provided with a version of the ComprehENotes test set that had been preprocessed with NoteAid. We preprocessed the ComprehENotes question text using NoteAid, extracted the simplifications and definitions that were provided, and used the NoteAid output as the question text shown to Turkers in the Treatment-NA group (refer to Figure 2 for an example of text simplified by NoteAid). The tasks were restricted so that individuals who completed 1 were not eligible to complete the other 2. For all groups, we collected demographic information about the Turkers' age, gender, ethnicity, level of education, and occupation. We also administered the S-TOFHLA test for each group to assess functional health literacy and to compare S-TOFHLA and ComprehENotes scores.

As we are not able to monitor the Turkers as they complete our tasks, we cannot know for sure that the baseline group did not use any external resources as instructed. However, we can be confident that they did not have access to NoteAid. To access NoteAid, the Turkers would have to have known the URL link to access the system, even though we did not provide it to them. Alternatively, the Turkers would have had to search for NoteAid without knowing the name of the specific system we are testing. Therefore, we are confident that even if the baseline group did use some external source during the task, they did not have access to NoteAid. The baseline Turkers may have found MLP if they searched on the Web for medical concepts during the task. For example, a Google search of "COPD definition" returns an MLP link on the first page. However, unless the Turkers knew about MLP before beginning the task, it is unlikely that they would use MLP as a reference during the task.

Figure 1. Flowchart describing our experiment. Amazon Mechanical Turk workers were randomly assigned to one of three tasks on the platform. They completed the ComprehENotes test with the use of the provided external tool. All scores were then collected, and ability estimated were obtained using Item Response Theory (IRT).



Textbox 1. Sample questions taken from the ComprehENotes test.

Instructions

Please read the following questions, making a note of the italicized text, and then examine the provided answer choices. Please select the answer that best represents the italicized portion of the question text.

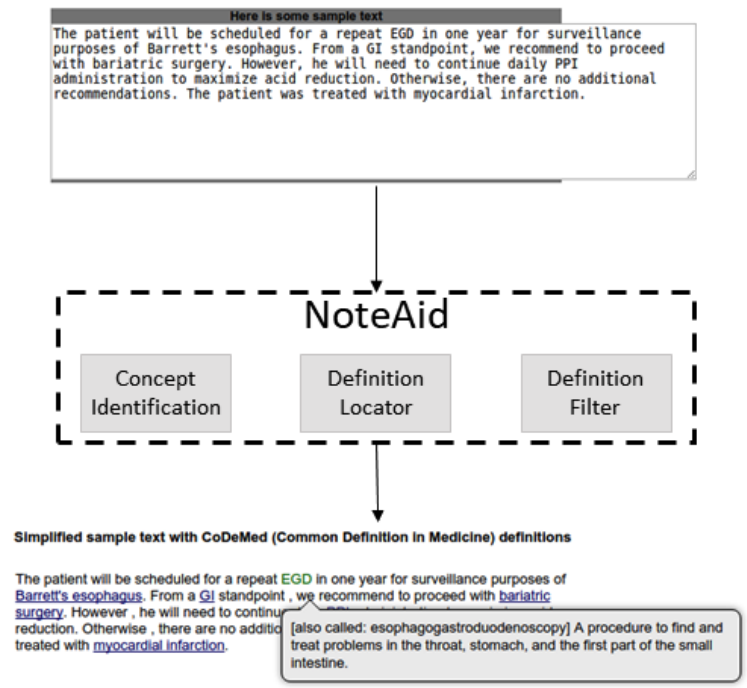
Amitriptyline 25 mg po at bedtime; Bactrim 160 mg po bid on Friday, Saturday, and Sunday; hydrocortisone cream; and *pegfilgrastim 6 mg subcutaneous one dose*. He will continue to return for scheduled chemotherapy and will also be following up with the hematology and oncology clinic.

1. Do a under skin injection of 1 dose of 6 mg pegfilgrastim.
2. Pegfilgrastim 6 mg epidermal 1 dose.
3. Pegfilgrastim may prevent neutropenia.

The patient is in for her physical examination today. Overall, she is doing very well. She is not on any blood pressure medications at the moment; she is doing fine. She had some issues in the past, but those settled down. Her blood pressure is 110/78 today on no medications, pulse 68 and regular, respirations 12.

1. No physical examination was performed for the patient today.
2. The patient came to check her health.
3. An eye exam is not a part of a regular physical examination.

Figure 2. Example showing NoteAid simplified text.



We included quality control checks for our AMT tasks to ensure a high-quality response from the Turkers. First, we restricted access to our tasks to Turkers with a prior approval rating above 95% to include only Turkers whose work has been judged as high quality by other requesters. We also restricted the task to Turkers located in the United States as a proxy for a test of English proficiency. Within the actual task, we included 3 quality-check questions, which consisted of a very simple question with an obvious answer. If any Turker answered 1 or more of the quality control checks incorrectly, their responses were removed from the later analyses.

NoteAid

The NoteAid system supplies lay definitions for medical concepts in EHR notes [18,19]. Users enter the text from their EHR notes into the NoteAid system, which outputs a version of the note with medical concepts defined. When the user hovers his or her mouse over a concept, a popup with the definition is shown. Figure 2 shows a high-level overview of the components in the NoteAid system, with example text that has been annotated. Users enter their EHR note text into NoteAid and are provided with a reproduction of the text, with key medical concepts linked to their definitions.

Figure 3. Equations for Item Response Theory 3-parameter logistic models.

$$p_i(\theta_j) = c_i + \frac{1 - c_i}{1 + e^{-a_i(\theta_j - b_i)}} \quad (1)$$

$$q_i(\theta_j) = 1 - p_{ij}(\theta_j) \quad (2)$$

$$p(U_j|\theta_j) = \prod_{i=1}^{I} p_i(\theta_j)^{u_{ij}} q_i(\theta_j)^{(1-u_{ij})} \quad (3)$$

NoteAid consists of 2 components. The *concept identifier* component processes input text and maps terms to medical concepts. The concepts are mapped to entries in the Unified Medical Language System using MetaMap [41,42]. It then filters the list of returned concepts to include only concepts that match a subset of possible semantic types related to patient health (eg, disease or syndrome and lab or test result). The *definition fetcher* component uses the filtered list of concepts to pull definitions from an external knowledge resource (eg, Wikipedia or MLP).

Previous evaluation of NoteAid has shown that patients' self-reported comprehension scores improve when using the system [18,19]. However, there has not yet been an evaluation of NoteAid on a test of comprehension, as opposed to self-reporting scores.

Item Response Theory Analysis

The ComprehENotes test set was developed using IRT [40]. The test set was built according to a single factor, 3-parameter logistic IRT model with a fixed guessing parameter. The test, therefore, measures a single latent trait, specifically the ability to comprehend EHR notes. Once the model has been fit, ability for a new test respondent is estimated by estimating θ according to the respondent's answers to the test questions after the responses have been converted to a correct or incorrect binary format. For a single test question *i*, the probability that individual *j* answers the question correctly is a function of the individual's ability (θ). Figure 3 includes 3 equations: equation 1 is used to calculate the probability that individual *j* with an estimated ability of θ_j will answer question *i* correctly; equation 2 calculates the probability that individual *j* with estimated ability θ_j will answer question *i* incorrectly; and equation 3 calculates the likelihood of individual *j* 's set of responses U_j to all items in the test set, where u_{ij} is 1 if individual *j* answered item *I* correctly and 0 if they did not.

p_i and q_i are functions of the known item parameters, and therefore, we can estimate θ via maximum likelihood for each Turker. We also calculated raw test scores for each Turker (percent of questions answered correctly) for comparison.

Results

Turker Demographics

We first report the demographic information for the Turkers who completed our tasks. Table 1 shows the demographic information that we collected from the Turkers for the Baseline, Treatment-MLP, and Treatment-NA groups. Overall, most of the Turkers who completed our tasks are white, young, and have at least an associate degree. In addition, most of the Turkers do not work in the medical field. These demographics are not representative of a wider population and do not fit demographics that are more commonly associated with low health literacy [31]. However, our goal here is to compare the results with respect to different interventions. In this case, we do not need to test individuals with low health literacy; we instead want to see if scores improve when users are provided with certain external resources.

Influence of Interventions

Our analysis includes both the raw test scores as well as the estimated ability level using IRT. As the test set consists of questions that were fit using IRT, we can also calculate the ability of these Turkers and test whether the mean ability score was higher for Turkers that used NoteAid. Ability is a useful metric as it takes into consideration which questions you answer correctly, not just how many. IRT models question difficulty, so by considering whether easy or difficult answers were correct, IRT allows for a more informative score than percent correct. For each Turker, we calculated their ability score (θ) using the IRT model fit as part of the ComprehENotes dataset [20]. We use the *mirt* and *ltm* open-source R packages for estimation [43,44].

Figure 4 plots the raw scores for each AMT Turker for our test set. The center rectangles span the range from the first quartile to the third quartile of responses, and the bolded line inside each box represents the median score. Open circles indicate outlier scores. The upper horizontal line marks the maximum score for each group, and the lower horizontal line is 1.5 times the interquartile range below the first quartile. As the figure shows, visually there is a spread between the populations that did and did not have access to the interventions. Median raw scores for the baseline and MLP groups are similar, whereas median scores

for the NoteAid group is higher. The spread of responses for the treatment groups is also smaller than the baseline group.

Figure 5 shows the box plots of ability estimates. Again, the median values for the baseline and MLP groups are similar and the median ability estimates for the NoteAid group is higher. The lowest ability estimates for the baseline and MLP groups are much lower than for the NoteAid group (2 SDs below the mean as opposed to 1 SD below). This shows that even for individuals that use NoteAid and still struggle, the low range of ability is higher than when NoteAid is not used.

To test whether either intervention caused a significant difference in scores, we compared each intervention with our baseline using Welch 2-sample *t* test. Table 2 shows the mean raw scores and mean ability estimates for Turkers in each group. Mean scores are significantly higher than the baseline for Turkers that had access to NoteAid, both with regard to the raw scores (P=.01) and estimated ability (P=.02).

Regression Analysis

We also wanted to determine if demographic factors had an impact on test scores. To that end, we fit a linear regression model to predict raw scores using demographic information and group (eg, baseline or treatment) as features. The results of the analysis showed that the intervention (none, MLP, or NoteAid) was a significant feature in predicting raw score. In addition, certain demographic groups were significant in determining score. Regarding ethnicity, individuals who self-reported as white had a significant positive coefficient. Regarding education, individuals that have less than a high school degree had a significant negative coefficient. These results are consistent with what is known about populations that are at risk for low health literacy. Individuals with lower education often have higher instances of low health literacy, as well as minorities. Our populations for this task, particularly with regard to minorities and less educated individuals, were very small. Future work on NoteAid in minority populations would be worthwhile to confirm these effects.

Table 1. Demographic information collected from Turkers who completed our task.

Demographic	Baseline (N=41), n (%)	MedlinePlus (N=29), n (%)	NoteAid (N=27), n (%)	Total (N=97), n (%)
Gender				
Male	27 (66)	8 (28)	18 (67)	53 (55)
Female	14 (34)	21 (72)	9 (33)	44 (45)
Age (years)				
22-34	23 (56)	16 (55)	16 (59)	55 (57)
35-44	6 (15)	9 (31)	8 (30)	23 (24)
45-54	8 (20)	2 (7)	3 (11)	13 (13)
55-64	4 (10)	2 (7)	0 (0)	6 (6)
65 and older	0 (0)	0 (0)	0 (0)	0 (0)
Ethnicity				
American Indian or Alaska Native	0 (0)	1 (3)	1 (4)	2 (2)
Asian	3 (7)	0 (0)	1 (4)	4 (4)
Black or African American	8 (20)	3 (10)	4 (15)	15 (16)
Hispanic	4 (10)	1 (3)	0 (0)	5 (5)
White	26 (63)	24 (83)	21 (78)	71 (73)
Education				
Less than high school	1 (2)	0 (0)	0 (0)	1 (1)
High school diploma	9 (22)	8 (28)	8 (30)	25 (26)
Associates	8 (20)	5 (17)	3 (11)	16 (17)
Bachelors	20 (49)	14 (48)	14 (51)	48 (50)
Masters or higher	3 (7)	2 (7)	2 (7)	7 (7)
Occupation				
Physician	0 (0)	0 (0)	1 (4)	1 (1)
Nurse	2 (5)	0 (0)	0 (0)	2 (2)
Medical student	1 (2)	1 (3)	1 (4)	3 (3)
Other profession in medicine	2 (5)	3 (10)	3 (11)	8 (8)
Other profession	36 (88)	25 (86)	22 (82)	83 (86)

Figure 4. Box plot of raw scores for baseline and treatment Turker groups. The treatment groups were able to use MedlinePlus and NoteAid, respectively, when taking the ComprehENotes test.

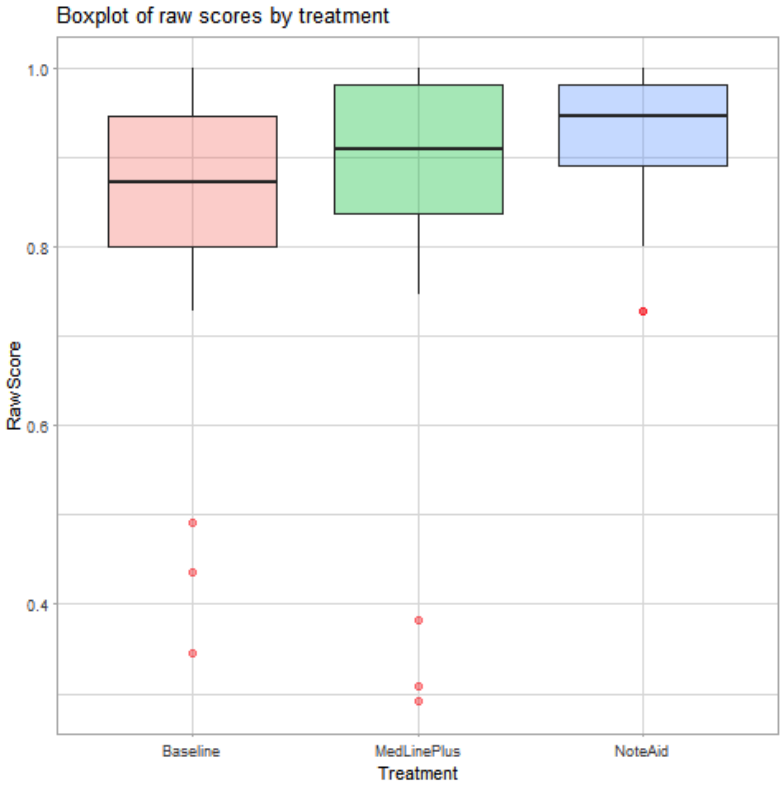


Figure 5. Box plot of ability estimates for baseline and treatment Turker groups. The treatment groups MLP and NA were able to use MedlinePlus and NoteAid, respectively, when taking the ComprehENotes test. IRT: Item Response Theory.

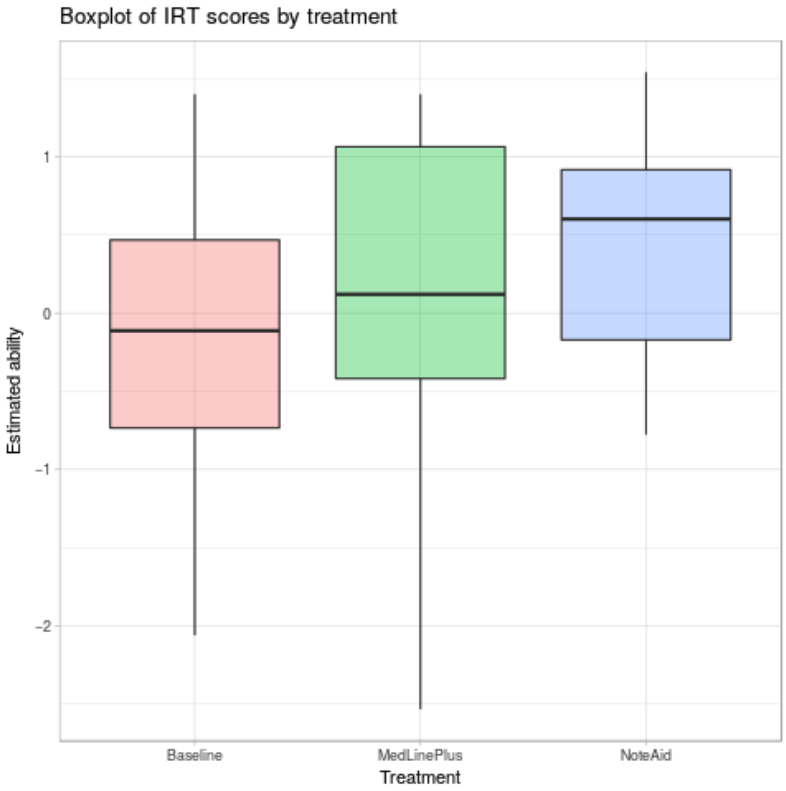


Table 2. Mean scores for the 3 groups. Mean NoteAid scores are significantly higher than the mean baseline scores, both for raw scores (*P*=.01) and estimated ability (*P*=.02).

Group	Raw score	Ability estimate
Baseline	0.831	−0.065
MedlinePlus	0.849	0.138
NoteAid	0.923[a]	0.477[a]

[a]Score significantly higher than baseline.

Comparison With the Short Test of Functional Health Literacy in Adults

All Turkers who completed our tasks were also given the S-TOFHLA test to complete. Scores on S-TOFHLA place test-takers into 1 of the 3 categories: *inadequate health literacy*, *marginal health literacy*, and *adequate health literacy*. It is most useful as a screening tool to identify individuals with low or marginal health literacy. All Turkers in our tasks were scored to have *adequate health literacy*. In fact, all Turkers either scored perfect scores or only answered 1 question incorrectly, whereas the scores from the ComprehENotes test covered a wide range of ability estimates. The ComprehENotes can be used to assess EHR note comprehension at a more granular level as opposed to a screening tool such as S-TOFHLA, where the primary concern is identification of individuals with low health literacy.

ComprehENotes Analysis

Finally, we wanted to see if the IRT model that was originally fit as part of the ComprehENotes dataset was validated by the response patterns that we collected from the Turkers. To this end, we selected the 2 questions that the most Turkers answered correctly as well as the 2 questions that the fewest Turkers answered correctly.

These questions can be considered the easiest and hardest, respectively, from our task. The difficulty parameters for these items as modeled by IRT match the expectation of how difficult these items should be. The 2 hardest questions from our task (in terms of how many Turkers answered correctly) have difficulty parameters of 0.7 and −0.3, whereas the 2 easiest questions have difficulty parameters of −1.8 and −1.4. The difficulty parameter is associated with the level of ability at which an individual has a 50% chance of answering the question correctly. Therefore, the low difficulty levels imply that someone of low ability has a 50% chance of answering the question correctly. Conversely, a higher difficulty parameter means that someone must be of a higher estimated ability level to have a 50% chance of answering correctly.

Discussion

Principal Findings

In this work, we have shown the importance of targeted, active intervention when trying to improve a person's ability to comprehend EHR notes. By giving Turkers access to NoteAid, scores on the ComprehENotes test are significantly improved over a baseline population that had no external resources. On the other hand, Turkers that had access to MLP but had to search themselves for the information that they wanted did not have a significant improvement in scores. NoteAid automatically identifies key medical concepts and provides definitions, as opposed to the scenario with MLP, where a user must decide what to search for. The user may not know that a certain concept is key for understanding a passage or they may assume that they understand certain concepts that they do not. By letting the user decide what to search for, important terms may be missed and overall comprehension may be affected. This result is consistent with previous work on assessing comprehension using tools such as NoteAid [18,19], but this is the first time where the conclusion is based on an EHR note comprehension assessment instead of patient self-reported scores. By using the ComprehENotes test, we can quantitatively confirm the previous results self-reported by patients.

Limitations

There are limitations to this work. First, by using AMT, we are not able to monitor the Turkers who complete our task to ensure that only the external resources that we provide were used. This is particularly true in the baseline group, where our expectation is that no external resource was used. However, it is unlikely that the baseline users were able to access NoteAid without prior knowledge of the system; therefore, we can be confident that they did not use it in our task. If the baseline users did use external resources, they most likely used a passive resource such as Google or even MLP. As NoteAid was integrated into the Treatment-NA task, we can be confident that Turkers in the Treatment-NA task used NoteAid. The discrepancy between Treatment-MLP and Treatment-NA may seem to bias improvements toward the Treatment-NA group, but there is an important distinction to be made. At present, sites such as MLP are available to any patient that seeks them out, but the onus is on the patient to go to the site and search for terms. With the Treatment-NA group, we have shown that by integrating a system that can simplify and define medical terms automatically, the burden of defining terms is removed from the patient.

In addition, the demographics of the Turkers who completed our task are not representative of the larger population, specifically among demographics associated with higher risks of low health literacy [31]. In the case of this work, that is not problematic, as our goal was to examine the effect of active and passive interventions on EHR note comprehension. The demographics of our 3 groups were similarly distributed, so the changes in scores can be linked to the intervention used. Although the results obtained were significant, ideally larger populations could be examined in each group. However, as the demographics of the Turkers are not consistent with demographic groups associated with low health literacy, the

follow-up work should focus on those groups. By using AMT and Turkers, we have shown that tools such as NoteAid do improve EHR note comprehension generally, but future work should look specifically at groups associated with low health literacy to determine if our results hold for those groups as well.

Another limitation of this study is that patients are not evaluated on their own notes. Ideally, we would be able to assess the EHR note comprehension of each patient by testing the patient using concepts extracted from his or her own EHR notes. However, there are several roadblocks to making this a reality. First, this type of personalized assessment would reduce the ability to compare comprehension ability between patients. If a patient scores highly on an assessment of their own note, we can say that the patient understands the note, but if there were no complex concepts in the note, we cannot compare this with a patient who scores poorly on an evaluation based on his or her own complex EHR note. Second, to build a personalized EHR note evaluation would require complex natural language processing (NLP) systems to automatically generate multiple-choice questions (MCQs) for patients when they enter their EHR notes. To our knowledge, there does not currently exist an NLP system for medical MCQ generation. We do believe that the development of such a system will be beneficial for personalized patient assessment of EHR note comprehension. Such a personalized system could complement the ComprehENotes test so that a patient would be assessed on their own EHR note as well as on a standardized assessment.

Conclusions

In this work, we have shown that simply having access to resources designed to improve health literacy and medical concept understanding is not enough to provide benefit. The Turkers in our experiment who had access to MLP did not score significantly higher on the ComprehENotes test than those Turkers that were not provided with an external resource. On the other hand, having access to NoteAid, which actively pulls definition information and provides it to the user, led to significantly higher scores for Turkers. This result validates previously reported self-scored comprehension results showing that users had an easier time understanding their notes when they had access to NoteAid.

Knowing that users do not see benefits from simply having access to MLP is an important observation. When doctors are recommending next steps for patients who wish to improve their health literacy, it may not be sufficient to point them to Web-based resources. Targeted interventions are necessary to ensure that patients are able to learn about specific concepts and diseases that are relevant to them. In particular, the integration of NoteAid with the EHR note on a patient's portal would remove the friction from the patient accessing an external resource. Instead, the patient would have key terms defined and simplified within his or her own patient portal, which would minimize the effort involved from the patient's standpoint and keep the information in the note within the portal itself.

There are several directions for future work. Developing target curricula is necessary to ensure that patients can see benefits from Web-based resources. They may not need a tool such as NoteAid (eg, if they are not looking at their notes), but something more targeted than MLP is needed to ensure that patients are learning. In addition, there should be further validation of the ComprehENotes test set with patients that are at risk for low health literacy. The Turkers in our task all scored either close to average or above average in our ability estimates, except for a few outliers. The test was designed to be most informative for individuals of lower ability, so this test should be replicated with such a population.

Acknowledgments

The authors thank the anonymous Turkers who participated in the Amazon Mechanical Turk tasks. This work was supported in part by the Investigator Initiated Research 1I01HX001457 from the Health Services Research and Development Program of the United States Department of Veterans Affairs (VA) and in part by a startup fund from the University of Massachusetts Medical School, both to HY. We also acknowledge the support of R01LM012817 from the National Institutes of Health (NIH). This work was also supported in part by the Center for Intelligent Information Retrieval (CIIR). The content is solely the responsibility of the authors and does not represent the views of CIIR, NIH, VA, the US Government, or the University of Massachusetts Medical School.

Conflicts of Interest

None declared.

References

1. Ross SE, Lin C. The effects of promoting patient access to medical records: a review. J Am Med Inform Assoc 2003 Apr;10(2):129-138 [FREE Full text] [Medline: 12595402]
2. Masys D, Baker D, Butros A, Cowles KE. Giving patients access to their medical records via the internet: the PCASSO experience. J Am Med Inform Assoc 2002;9(2):181-191 [FREE Full text] [Medline: 11861633]
3. Sheldon MG. Giving patients a copy of their computer medical record. J R Coll Gen Pract 1982 Feb;32(235):80-86 [FREE Full text] [Medline: 7097630]
4. Bronson DL, Costanza MC, Tufo HM. Using medical records for older patient education in ambulatory practice. Med Care 1986 Apr;24(4):332-339. [Medline: 3515060]

5. Homer CS, Davis GK, Everitt LS. The introduction of a woman-held record into a hospital antenatal clinic: the bring your own records study. Aust N Z J Obstet Gynaecol 1999 Feb;39(1):54-57. [Medline: 10099751]
6. Draper J, Field S, Thomas H, Hare MJ. Should women carry their antenatal records? Br Med J (Clin Res Ed) 1986 Mar 01;292(6520):603 [FREE Full text] [Medline: 3081189]
7. Hertz CG, Bernheim JW, Perloff TN. Patient participation in the problem-oriented system: a health care plan. Med Care 1976 Jan;14(1):77-79. [Medline: 1084944]
8. Baldry M, Cheal C, Fisher B, Gillett M, Huet V. Giving patients their own records in general practice: experience of patients and staff. Br Med J (Clin Res Ed) 1986 Mar 1;292(6520):596-598 [FREE Full text] [Medline: 3081187]
9. Golodetz A, Ruess J, Milhous RL. The right to know: giving the patient his medical record. Arch Phys Med Rehabil 1976 Feb;57(2):78-81. [Medline: 1083223]
10. Honeyman A, Cox B, Fisher B. Potential impacts of patient access to their electronic care records. Inform Prim Care 2005;13(1):55-60 [FREE Full text] [Medline: 15949176]
11. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
12. Jones RB, McGhee SM, McGhee D. Patient on-line access to medical records in general practice. Health Bull (Edinb) 1992 Mar;50(2):143-150. [Medline: 1517087]
13. US National Library of Medicine. Fact Sheet: MedlinePlus® URL: http://wayback.archive-it.org/org-350/20180312141606/https://www.nlm.nih.gov/pubs/factsheets/medlineplus.html [accessed 2018-12-03] [WebCite Cache ID 74NocOqTv]
14. Coberly E, Boren SA, Davis JW, McConnell AL, Chitima-Matsiga R, Ge B, et al. Linking clinic patients to Internet-based, condition-specific information prescriptions. J Med Libr Assoc 2010 Apr;98(2):160-164 [FREE Full text] [doi: 10.3163/1536-5050.98.2.009] [Medline: 20428282]
15. Gaines JK, Levy LS, Cogdill KW. Sharing MedlinePlus®/MEDLINE® for information literacy education (SMILE): a dental public health information project. Med Ref Serv Q 2011;30(4):357-364 [FREE Full text] [doi: 10.1080/02763869.2011.609016] [Medline: 22040242]
16. McMullen KD, McConnaughy RP, Riley RA. Outreach to improve patient education at South Carolina free medical clinics. J Consum Health Internet 2011;15(2):117-131 [FREE Full text] [doi: 10.1080/15398285.2011.572779] [Medline: 22084623]
17. Teolis MG. A MedlinePlus kiosk promoting health literacy. J Consum Health Internet 2010;14(2):126-137 [FREE Full text] [doi: 10.1080/15398281003780966] [Medline: 20808715]
18. Polepalli RB, Houston T, Brandt C, Fang H, Yu H. Improving patients' electronic health record comprehension with NoteAid. Stud Health Technol Inform 2013;192:714-718. [Medline: 23920650]
19. Polepalli RB, Yu H. Systems for Improving Electronic Health Record Note Comprehension. 2013 Presented at: ACM SIGIR Workshop on Health Search & Discovery; 2013; Dublin.
20. Lalor JP, Wu H, Chen L, Mazor KM, Yu H. ComprehENotes, an instrument to assess patient reading comprehension of electronic health record notes: development and validation. J Med Internet Res 2018 Apr 25;20(4):e139 [FREE Full text] [doi: 10.2196/jmir.9380] [Medline: 29695372]
21. Snow R, O'Connor B, Jurafsky D, Ng A. Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. USA: Association for Computational Linguistics; 2008 Presented at: EMNLP '08; October 25 - 27, 2008; Honolulu, Hawaii URL: http://dl.acm.org/citation.cfm?id=1613715.1613751
22. Sabou M, Bontcheva K, Scharl A. Crowdsourcing Research Opportunities: Lessons from Natural Language Processing. USA: ACM; 2012 Presented at: International Conference on Knowledge Management and Knowledge Technologies; September 05 - 07, 2012; Graz, Austria. [doi: 10.1145/2362456.2362479]
23. Lalor JP, Wu H, Yu H. Building an evaluation scale using item response theory. Proc Conf Empir Methods Nat Lang Process 2016 Nov;2016:648-657 [FREE Full text] [Medline: 28004039]
24. Demartini G, Difallah D, Cudré-Mauroux P. ZenCrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. USA: ACM; 2012 Presented at: International Conference on World Wide Web; April 16 - 20, 2012; Lyon, France. [doi: 10.1145/2187836.2187900]
25. Lalor J, Wu H, Munkhdalai T, Yu H. Understanding Deep Learning Performance through an Examination of Test Set Difficulty: A Psychometric Case Study. 2018 Presented at: Empirical Methods in Natural Language Processing; October 31–November 4; Brussels, Belgium.
26. Zhai H, Lingren T, Deleger L, Li Q, Kaiser M, Stoutenborough L, et al. Web 2.0-based crowdsourcing for high-quality gold standard development in clinical natural language processing. J Med Internet Res 2013;15(4):e73 [FREE Full text] [doi: 10.2196/jmir.2426] [Medline: 23548263]
27. Good BM, Nanis M, Wu C, Su AI. Microtask crowdsourcing for disease mention annotation in PubMed abstracts. Pac Symp Biocomput 2015:282-293 [FREE Full text] [Medline: 25592589]
28. Mortensen JM, Minty EP, Januszyk M, Sweeney TE, Rector AL, Noy NF, et al. Using the wisdom of the crowds to find critical errors in biomedical ontologies: a study of SNOMED CT. J Am Med Inform Assoc 2015 May;22(3):640-648 [FREE Full text] [doi: 10.1136/amiajnl-2014-002901] [Medline: 25342179]

29. Gottlieb A, Hoehndorf R, Dumontier M, Altman R. Ranking adverse drug reactions with crowdsourcing. J Med Internet Res 2015 Mar 23;17(3):e80 [FREE Full text] [doi: 10.2196/jmir.3962] [Medline: 25800813]
30. Kutner M. Eric. 2006. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy. NCES 2006-483 URL: https://eric.ed.gov/?id=ED493284 [accessed 2018-12-03] [WebCite Cache ID 74Np3gMga]
31. Lynn N, Allison M. Health literacy: a prescription to end confusion. Washington (DC): The National Academies Press; 2004.
32. Lerner EB, Jehle DV, Janicke DM, Moscati RM. Medical communication: do our patients understand? Am J Emerg Med 2000 Nov;18(7):764-766. [doi: 10.1053/ajem.2000.18040] [Medline: 11103725]
33. Chapman K, Abraham C, Jenkins V, Fallowfield L. Lay understanding of terms used in cancer consultations. Psychooncology 2003 Sep;12(6):557-566. [doi: 10.1002/pon.673] [Medline: 12923796]
34. Halbach SM, Enders A, Kowalski C, Pförtner T, Pfaff H, Wesselmann S, et al. Health literacy and fear of cancer progression in elderly women newly diagnosed with breast cancer--a longitudinal analysis. Patient Educ Couns 2016 May;99(5):855-862. [doi: 10.1016/j.pec.2015.12.012] [Medline: 26742608]
35. Reading SR, Go AS, Fang MC, Singer DE, Liu IA, Black MH, Anticoagulation and Risk Factors in Atrial Fibrillation–Cardiovascular Research Network (ATRIA-CVRN) Investigators. Health literacy and awareness of atrial fibrillation. J Am Heart Assoc 2017 Apr 11;6(4):e005128 [FREE Full text] [doi: 10.1161/JAHA.116.005128] [Medline: 28400367]
36. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med 1995 Oct;10(10):537-541. [Medline: 8576769]
37. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999 Sep;38(1):33-42. [Medline: 14528569]
38. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med 2005;3(6):514-522 [FREE Full text] [doi: 10.1370/afm.405] [Medline: 16338915]
39. Royer JM, Hastings CN, Hook C. A sentence verification technique for measuring reading comprehension. J Lit Res 2016 Sep 11;11(4):355-363. [doi: 10.1080/10862967909547341]
40. Baker F, Kim SH. Item Response Theory: Parameter Estimation Techniques, Second Edition. Boca Raton, Florida: CRC Press; 2004.
41. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001:17-21 [FREE Full text] [Medline: 11825149]
42. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267-D270 [FREE Full text] [doi: 10.1093/nar/gkh061] [Medline: 14681409]
43. Rizopoulos D. ltm: an R package for latent variable modeling and item response analysis. J Stat Soft 2006;17(5):1-25. [doi: 10.18637/jss.v017.i05]
44. Chalmers RP. mirt: a multidimensional item response theory package for the R environment. J Stat Soft 2012;48(6):1-29 [FREE Full text] [doi: 10.18637/jss.v048.i06]

Abbreviations

AMT: Amazon Mechanical Turk
COPD: chronic obstructive pulmonary disease
EHR: electronic health record
IRT: Item Response Theory
MCQ: multiple-choice question
MLP: MedlinePlus
NLP: natural language processing
NVS: Newest Vital Sign
S-TOFHLA: short Test of Functional Health Literacy in Adults
TOFHLA: Test of Functional Health Literacy in Adults
Treatment-NA: Treatment-NoteAid

Edited by G Eysenbach; submitted 27.04.18; peer-reviewed by S Brunak, J Jones; comments to author 19.07.18; revised version received 28.09.18; accepted 26.10.18; published 16.01.19.

Please cite as:
Lalor JP, Woolf B, Yu H
Improving Electronic Health Record Note Comprehension With NoteAid: Randomized Trial of Electronic Health Record Note Comprehension Interventions With Crowdsourced Workers
J Med Internet Res 2019;21(1):e10793
URL: http://www.jmir.org/2019/1/e10793/
doi:10.2196/10793
PMID:30664453

Original Paper

EuroQol (EQ-5D-5L) Validity in Assessing the Quality of Life in Adults With Asthma: Cross-Sectional Study

Gimena Hernandez[1,2,3], MD, MPH; Olatz Garin[1,3,4], MPH, PhD; Alexandra L Dima[5], PhD; Angels Pont[1,3], BSc; Marc Martí Pastor[1,2,3], MD, MPH; Jordi Alonso[1,3,4], MD, PhD; Eric Van Ganse[5,6], MD, PhD; Laurent Laforest[6], MD, PhD; Marijn de Bruin[7], PhD; Karina Mayoral[1,2], MPH; Vicky Serra-Sutton[3,8], PhD; Montse Ferrer[1,2,3], MD, PhD; ASTRO-LAB Group

[1]Health Services Research Group, Hospital del Mar Medical Research Institute, Barcelona, Spain

[2]Department of Paediatrics, Obstetrics and Gynaecology and Preventive Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain

[3]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública, Madrid, Spain

[4]Experimental and Health Sciences, Pompeu Fabra University, Barcelona, Spain

[5]Health Services and Performance Research, Universite´ Claude Bernard Lyon 1, Lyon, France

[6]Pharmaco-Epidemiology Lyon, Lyon, France

[7]Institute of Applied Health Sciences, University of Aberdeen, Scotland, United Kingdom

[8]Agència de Qualitat i Avaluació Sanitàries de Catalunya, Barcelona, Spain

Corresponding Author:
Montse Ferrer, MD, PhD
Health Services Research Group
Hospital del Mar Medical Research Institute
Dr Aiguader 88
Barcelona, 08003
Spain
Phone: 34 933160763
Fax: 34 933160797
Email: mferrer@imim.es

Abstract

Background: The EuroQol-5 Dimension (EQ-5D), developed in 1990, is a most widely used generic tool to measure the health-related quality of life (HRQoL) and considered suitable for patients with asthma. In 2009, the EuroQol Group developed a new EQ-5D version to overcome limitations related to its consistently reported high ceiling effect. To enhance the sensitivity for assessing the HRQoL in further patient populations, the number of responses of EQ-5D was increased from 3 to 5 levels (EQ-5D-5L). Moreover, the availability of well-defined requirements for its Web-based administration allows EQ-5D-5L use to monitor the HRQoL in electronic health (eHealth) programs. No study has evaluated the metric properties of the new EQ-5D-5L in patients with asthma yet.

Objective: This study aims to examine the distribution, construct validity, and reliability of the new EQ-5D-5L questionnaire administered online to adults with asthma.

Methods: We evaluated patients with asthma (age: 18-40 years) from a primary care setting in France and England, who self-completed the EQ-5D-5L questionnaire online. The inclusion criteria were persistent asthma defined as >6 months of prescribed inhaled corticosteroids and long-acting beta-agonists or inhaled corticosteroids alone during the 12 months prior to inclusion. The EQ-5D index was obtained by applying the English preference value set for the new EQ-5D-5L and the French 3L-5L crosswalk value set. Both value sets produced single preference-based indices ranging from 1 (best health state) to negative values (health states valued as worse than death), where 0=death, allowing the calculation of quality-adjusted life years. Responses to dimensions and index distribution, including ceiling and floor effects, were examined. The construct validity was assessed by comparing the means of known groups by analyses of variance and calculation of effect sizes.

Results: Of 312 patients answering the baseline Web-based survey, 290 completed the EQ-5D-5L (93%). The floor effect was null, and the ceiling effect was 26.5% (74/279). The mean EQ-5D-5L index was 0.88 (SD 0.14) with the English value set and 0.83 (SD 0.19) with the French 3L-5L crosswalk value set. In both indices, large effect sizes were observed for known groups defined by the Asthma Control Questionnaire (1.06 and 1.04, P<.001). Differences between extreme groups defined by chronic

conditions (*P*=.002 and *P*=.003 for the English value set and French 3L-5L crosswalk value set, respectively), short-acting beta-agonists (SABAs) canisters in the last 12 months (*P*=.02 and *P*=.03), or SABA use during the previous 4 weeks (*P*=.03 and *P*=.01) were of moderate magnitude with effect sizes around 0.5.

Conclusions: The new EQ-5D-5L questionnaire has an acceptable ceiling effect, a good construct validity based on the discriminant ability for distinguishing among health-related known groups, and high reliability, supporting its adequacy for assessing the HRQoL in patients with asthma. EQ-5D-5L completion by most Web-based respondents supports the feasibility of this administration form.

(J Med Internet Res 2019;21(1):e10178) doi:10.2196/10178

KEYWORDS

asthma; EQ-5D-5L; EuroQol; health-related quality of life; Web-based survey; validity; patient-reported outcome measures

Introduction

The impact of asthma on the patients' health has been traditionally assessed by either clinical markers or functional tests [1]. Patient-reported outcome measures, such as symptom control or health-related quality of life (HRQoL), have shown to be useful for clinical management, understanding disease impact on the patients' functional status and well-being, and cost-effectiveness analyses [2]. Hence, international guidelines for asthma diagnosis and treatment have emphasized that treatment goals should include the improvement of the patients' HRQoL [3].

In asthma, disease-specific HRQoL measures have been more widely used than generic ones, as they could be more sensitive. Adding generic HRQoL domains important to patients with asthma has been proposed [4] because asthma-specific HRQoL instruments measure similar contents to those covered by asthma control questionnaires [5,6] such as symptoms and activity limitations. Generic HRQoL instruments are broad measures that can be applied in patients with various conditions and the general population. The EuroQol-5 Dimensions (EQ-5D), developed in 1990 by the EuroQol Group, is one of the most widely used generic tools owing to its low respondent burden, good psychometric properties, and econometric development [7-9]. In addition, the availability of well-defined requirements for its Web-based administration by multiple devices, such as personal computer, tablet, or smartphones, makes this instrument adequate for monitoring the HRQoL in eHealth programs [10].

The EQ-5D was considered a suitable generic measure in a systematic review [11] of patient-reported outcome measures for patients with asthma. This health status measure allows the calculation of quality-adjusted life years (QALYs) when society preferences are applied and cost-utility analysis in economic evaluations [12-14]. However, to the best of our knowledge, only 3 studies have evaluated its psychometric properties in patients with asthma [15-17]. Garratt et al [16] showed a moderate EQ-5D association with asthma-specific HRQoL instruments and external variables such as smoking status and education level. Oga et al [15] and McTaggart-Cowan et al [17] reported a high ceiling effect (59% and 50% of the sample with the maximum score, respectively) questioning the usefulness of the EQ-5D in asthmatic patients. In fact, limitations related to the high ceiling effect have also been consistently reported for the EQ-5D in other chronic conditions such as chronic obstructive pulmonary disease [18], osteoarthritis [18], diabetes [19], and coronary heart disease [20].

The traditional EQ-5D descriptive system, composed of 5 dimensions with 3 levels of severity, defines 243 distinct health states resulting from all the possible combinations (ie, 3^5); this is a low number compared with other generic preference-based instruments such as the Health Utilities Index [21] or the SF-6D [22] with 972,000 and 18,000 possible combinations, respectively. To improve its sensitivity, the EuroQol Group developed a new EQ-5D version, by increasing the number of responses from 3 to 5 levels, known as EQ-5D-5L, with 3125 health states (ie, 5^5) [23].

The new EQ-5D-5L has already been tested in some disease-specific samples, such as patients with cancer [24,25] and with hepatitis [26], showing a better discrimination capability and lower ceiling effects than the traditional 3-level version (11% vs 17% [24], 9.7% vs 16.8% [25], and 21.6% vs 38.3% [26]). However, to date, no study has evaluated metric properties of the new 5-level EQ-5D in patients with asthma. Hence, this study aims to examine the distribution, construct validity, and reliability of the new EQ-5D-5L administered online to adults with asthma.

Methods

Setting and Study Population

In this study, we analyzed baseline data of adults (age: 18-40 years) enrolled in the ASTRO-LAB cohort who completed the EQ-5D-5L questionnaire. The ASTRO-LAB project is a prospective longitudinal study of asthmatic patients designed to provide new evidence regarding the safety of long-acting beta-agonists (LABAs) in routine primary care in France and the United Kingdom. Details of the study are described elsewhere [27].

The inclusion criteria were as follows: persistent asthma and age <40 years. Patients were considered to have persistent asthma when they had >6 months of prescribed treatment with inhaled corticosteroids (ICs) and LABAs or ICs alone during the 12 months prior to inclusion. Persistent asthma requires controller therapy on a regular basis, whereas intermittent asthma can be treated with rescue medication as needed. The ASTRO-LAB persistent asthma definition was based on a minimal prescription duration level of antiasthmatic drugs because this method is considered less biased than the

practitioner's classification of asthma, and it is frequently used in database studies [28]. The ASTRO-LAB project's age limit was chosen to minimize the recruitment of patients with other comorbid conditions frequent at older ages, most importantly chronic obstructive pulmonary disease, often overlapped and difficult to exclude without specific tests.

The exclusion criteria were as follows: chronic oral corticosteroid use (≥15 consecutive days 3 months before inclusion), history of omalizumab therapy, and any other concomitant chronic respiratory disease (chronic obstructive pulmonary disease, cystic fibrosis, pulmonary fibrosis, bronchiectasis, or tuberculosis). Owing to ASTRO-LAB's main focus being LABAs safety, the abovementioned criteria based on the administration of other medications aimed at avoiding confounding with their adverse effects, implying that most patients with severe persistent asthma were excluded.

The ASTRO-LAB study has been approved by the Ethics and Regulatory Boards in France and the United Kingdom and was conducted in accordance with the Declaration of the World Medical Association. In France, approval was obtained from CCTIRS (Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine dela santé) on November 21, 2012 (Dossier N°12702), and the authorization from Commission Nationale d'Informatique et Liberté was obtained in May 17, 2013 (DR-2013-264). In the United Kingdom, according to the UK Research Governance Framework, the study was submitted to The West London Research Ethics Committee (REC), and the final approval was obtained on April 15, 2013 (REC Reference 12/LO/20139). Following the UK regulatory process, the ASTRO-LAB consortium submitted the protocol to the National Institute for Health Research Clinical Research Network (NIHR CRN) to launch the review by Primary Care Trust local sites. The first local approval was granted by the West London Primary Care Consortium on May 22, 2013. Informed consent was obtained from all participants prior to inclusion.

Measurement Instruments

Clinical data were extracted from medical records, and patient-reported information was obtained by the following 2 administration modes: patient-completed Web-based survey and telephone interviews with patients performed by trained interviewers. The EQ-5D-5L was only administered in the Web-based survey.

Clinical Data

Information on age, gender, body mass index, comorbidity, and treatment prescribed was obtained; in France, general practitioners completed a Web-based survey at patient inclusion, while in the United Kingdom, this information was directly extracted from medical records. The history of 4 associated pathologies (allergic rhinitis, nasal polyps, anxiety or depression, and gastroesophageal reflux), was registered and transformed into a count variable. The total number of short-acting beta-agonist (SABA) canisters prescribed in the 12 months prior to inclusion was transformed into a variable of 3 categories—0, 1-4, and ≥5 canisters.

Patient-Completed Web-Based Survey

Patients received instructions during the recruitment contact to self-complete a Web-based survey, which included the EQ-5D-5L to measure the HRQoL and sociodemographic data, such as their highest level of education and current work situation, among others.

The EQ-5D-5L is a brief, multiattribute, generic, health status measure composed of 5 questions with Likert response options (descriptive system) and a visual analog scale (EQ-VAS). The latter asks patients to rate their own health from 0 to 100 (the worst and best imaginable health, respectively). The descriptive system covers 5 dimensions of health (mobility, self-care, usual activities, pain or discomfort, and anxiety or depression) with 5 levels of severity in each dimension (no problems, slight problems, moderate problems, severe problems, and unable to perform or extreme problems).

Preference value sets used to obtain the index of the EQ-5D-5L were the 3L-5L crosswalk from the French 3L version [29], and the new EQ-5D-5L value set from England [30]. In both cases, single preference-based indices were produced ranging from 1 (the best health state) to negative values (health states valued as worse than death), where 0=death. The minimal important difference for the EQ-5D index was estimated as 0.07 [31].

Telephone Interviews

The telephonic interviews were computer-assisted to standardize the process. Trained interviewers administered questions to patients about their asthma control and treatment use, among others. Asthma control is defined as the extent to which the manifestations of asthma can be observed in a patient, or have been reduced or removed by treatment [32,33]; it reflects the suitability of the asthma treatment.

The Asthma Control Questionnaire (ACQ) is composed of 7 items—the top scoring 5 symptoms, FEV_1% predicted, and daily rescue bronchodilator use. A shorter version called ACQ-symptoms only [34] was developed to use when it is not feasible to collect data about the last 2 items, as in ASTRO-LAB. It assesses the frequency of the 5 asthma symptoms during the previous week on a 7-point Likert scale (0=no impairment, 6=maximum impairment). The overall score, calculated as the mean of item responses, ranges from 0 to 6. A score <0.75 was defined as well-controlled asthma; 0.75-1.5 as intermediate control; and >1.5 as not well-controlled asthma [35]. The results generated by the short versions have shown to be very similar to those of the complete ACQ, as well as its measurement properties (reliability, responsiveness, internal consistency, construct validity, and interpretability) [34].

The following question was asked to patients with SABA therapy prescription: "How often have you usually taken your 'reliever medication' (brand name) in the past 4 weeks? Every day; almost every day; once or twice every week; less than once a week; or I don't know."

Analytic Strategy

Sample characteristics were described by calculating percentages, or mean (SD) values, according to the variable type (detailed in tables and figures). To examine the nonresponse

bias, subjects who completed the Web-based survey were compared with those who had not completed this survey by a t test and chi-square test.

We calculated the percentages of responses to each EQ-5D-5L dimension. To examine the distribution of the EQ-5D index, we calculated statistics of central tendency, dispersion, asymmetry, and tail extremity, as well as the proportion and 95% CI of the individuals in the best possible (ceiling) and the worst possible (floor) health states [36]. To assess the reliability based on the internal consistency, the Cronbach alpha coefficient was estimated.

The construct validity examines whether the instrument adequately assesses the concept that it intends to measure [37], in this case the HRQoL. The strategy to evaluate the construct validity based on known groups consists of testing the ability of the instrument to discriminate among groups previously hypothesized as differing in the concept measured. The following variables were chosen to test the instrument's capacity to discriminate, as differences have been consistently shown in the HRQoL among groups defined by them [1,15,38,39]—the number of chronic conditions (as an indicator of general health), number of SABA canisters prescribed in the previous year, frequency of SABA inhaler use during the previous 4 weeks, and ACQ scores (as 3 indicators of asthma control). We hypothesized that asthma patients with worse general health or less asthma control report worse HRQoL.

To evaluate the discriminative capacity of the EQ-5D index and EQ-VAS among the known groups mentioned above, mean scores were compared using one-way analysis of variance and the Tukey studentized range (honestly significant difference) test for post-hoc comparisons; alpha was set at.05. To assess the magnitude of the differences Cohen effect sizes were calculated. General guidelines define an effect size of 0.2 as small, 0.5 as moderate, and 0.8 as large [40]. All analyses were conducted using the statistical package SPSS (IBM SPSS Statistics for Windows, Version 23.0, IBM Corp)

Results

Study Sample

Of 581 subjects with asthma aged 18-40 years composing the ASTRO-LAB cohort, 312 filled in the baseline Web-based survey (Web-based participation rate, 53.7%), but 22 of these did not complete the EQ-5D-5L (questionnaire nonresponse rate, 7.0%). Of 290 who fulfilled the EQ-5D-5L, 11 were excluded because they had missing data on all the variables selected to define known groups; hence, 279 patients were finally included in this analysis. Table 1 shows patients' baseline characteristics, comparing the included subjects with excluded ones (mainly because of not responding to the Web-based survey). Most of the included subjects were from France and had been treated with fixed-dose combinations of LABA and IC. More than half of them had completed a bachelor degree (66.9%, 184/275), and 72.6% (201/277) were employed in their usual jobs. These 2 variables were only available for patients included in the analysis, as they were recorded in the Web-based questionnaire. Nonrespondents were younger (aged 29.8 vs 31.0 years, P=.03), and presented higher ACQ mean scores (worse control) in comparison to respondents but did not differ in body mass index, treatment, number of other chronic conditions, SABA canisters prescribed last year, and frequency of SABA used in the previous 4 weeks.

5-Level EuroQoL-5 Dimension Version Distribution

Figure 1 shows the percentages of responses to each EQ-5D-5L dimension. Most subjects reported "no problems" in mobility (81.0%, 226/279) and self-care (98.2%, 274/279) dimensions, while only around half of the subjects endorsed this category in pain or discomfort (45.5%, 127/279) and anxiety or depression (48.0%, 134/279) dimensions. The "extreme problems" category was endorsed by 1 subject for pain and 7 for anxiety or depression.

Table 2 shows the distribution characteristics of EQ-5D-5L indices. In our sample, the EQ-5D-5L index constructed with the English value set ranged from 0.16 to 1 and from –0.074 to 1 when constructed with the French 3L-5L crosswalk value set. The mean was 0.88 (SD 0.14) for the English index and 0.83 (SD 0.19) for the French one. The Kurtosis statistics of 5.62 and 3.26, with skewness of –2.06 and –1.63, indicated that the asymmetry to the right part of the distribution and the tail extremity were greater in the index constructed with the English EQ-5D-5L value set. The floor effect was null, and the ceiling effect was 26.5% (74/279). Cronbach alpha coefficient was .69, achieving the recommended standard [36,37].

5-level EuroQoL-5 Dimension Version Construct Validity

Table 3 presents results on the construct validity of the EQ-5D-5L based on known groups. Both EQ-5D-5L indices showed statistically significant different means for all known groups evaluated, while EQ-VAS only showed statistically significant differences among groups defined by ACQ scores. The mean EQ-5D-5L index for asthmatic patients decreased significantly with an increase in the number of other chronic conditions from 0.91 to 0.82 with the English value set and from 0.86 to 0.75 with the French 3L-5L crosswalk. The effect size between patients with none and those with ≥2 other chronic conditions were 0.62 and 0.60 (moderate) with EQ-5D-5L indices. In addition, effect sizes were moderate between extreme groups defined by SABA canisters prescribed in the previous year (0.58 and 0.46), and by the SABA frequency during the last 4 weeks (both 0.5). Finally, among groups defined by ACQ scores, the effect size between well-controlled and intermediately controlled asthma was moderate (0.44 and 0.47) and large between well- and not well-controlled asthma (1.06 and 1.04).

Table 1. The characteristics of included and excluded subjects.

Characteristics	Included patients (n=279)	Excluded patients (n=302)	*P* value
Age (years), mean (SD)	31.0 (6.7)	29.8 (6.7)	.03
<25, n (%)	62 (22.2)	85 (28.1)	.10
25-34, n (%)	119 (42.7)	133 (44.0)	
≥35, n (%)	98 (35.1)	84 (27.8)	
Gender, n (%)			.47
Male	110 (39.4)	128 (42.4)	
Female	169 (60.6)	174 (57.6)	
Country			.01
France, n (%)	222 (79.6)	264 (87.7)	
United Kingdom, n (%)	57 (20.4)	37 (12.3)	
Missing (n)	0	1	
Body mass index (kg/m^2), mean (SD)	25.2 (6.2)	25.4 (5.8)	.79
Missing (n)	127	107	
Treatment with, n (%)			.18
LABA[a]	11 (3.9)	9 (3.0)	
ICs[b]	71 (25.4)	60 (19.9)	
LABA+ICs in separate inhalers	37 (13.3)	33 (10.9)	
Fixed LABA and ICs combination	160 (57.3)	200 (66.2)	
Other chronic conditions			.51
0 conditions, n (%)	66 (41.5)	80 (39.2)	
1 condition, n (%)	62 (39.0)	91 (44.6)	
≥2 conditions, n (%)	31 (19.5)	33 (16.2)	
Missing (n)	120	98	
Number of SABA[c] canisters prescribed (last year)			.75
0 canisters, n (%)	119 (53.6)	133 (50.2)	
1-4 canisters, n (%)	78 (35.1)	100 (37.7)	
≥5 canisters, n (%)	25 (11.3)	32 (12.1)	
Missing (n)	57	—[d]	
Frequency of SABA use reported by patient (last 4 weeks)			.63
Less than once a week, n (%)	166 (61.9)	171 (65.5)	
Once or twice every week, n (%)	71 (26.5)	60 (23.0)	
Almost every day or every day, n (%)	31 (11.6)	30 (11.5)	
Missing (n)	11	41	
Asthma Control Questionnaire, mean (SD)	1.01 (0.92)	1.35 (1.01)	<.001
Well-controlled (<0.75), n (%)	119 (44.6)	89 (34.1)	<.001
Intermediate (0.75-1.5), n (%)	82 (30.7)	63 (24.1)	
Not well-controlled (>1.5), n (%)	66 (24.7)	109 (41.8)	
Missing (n)	12	41	
Highest education			Not calculated
Secondary school or less, n (%)	13 (4.7)	—	
Sixth form or college, n (%)	41 (14.9)	—	

Characteristics	Included patients (n=279)	Excluded patients (n=302)	*P* value
Bachelor degree, n (%)	184 (66.9)	—	
Postgraduate, n (%)	37 (13.5)	—	
Missing (n)	4	—	
Work status			Not calculated
Employed at usual job, n (%)	201 (72.6)	—	
On light duty or restricted work, n (%)	1 (0.4)	—	
Paid leave or sick leave, n (%)	4 (1.4)	—	
Unemployed because of other reason, n (%)	23 (8.3)	—	
Student (school, college, university), n (%)	35 (12.6)	—	
Keeping house or homemaker, n (%)	7 (2.5)	—	
Retired, n (%)	0 (0.0)	—	
On disability, n (%)	6 (2.2)	—	
Missing (n)	2	—	

[a]LABA: long-acting beta-agonist.

[b]IC: inhaled corticosteroid.

[c]SABA: short-acting beta-agonist.

[d]Indicates missing data.

Figure 1. The percentage of patients' responses to each dimension.

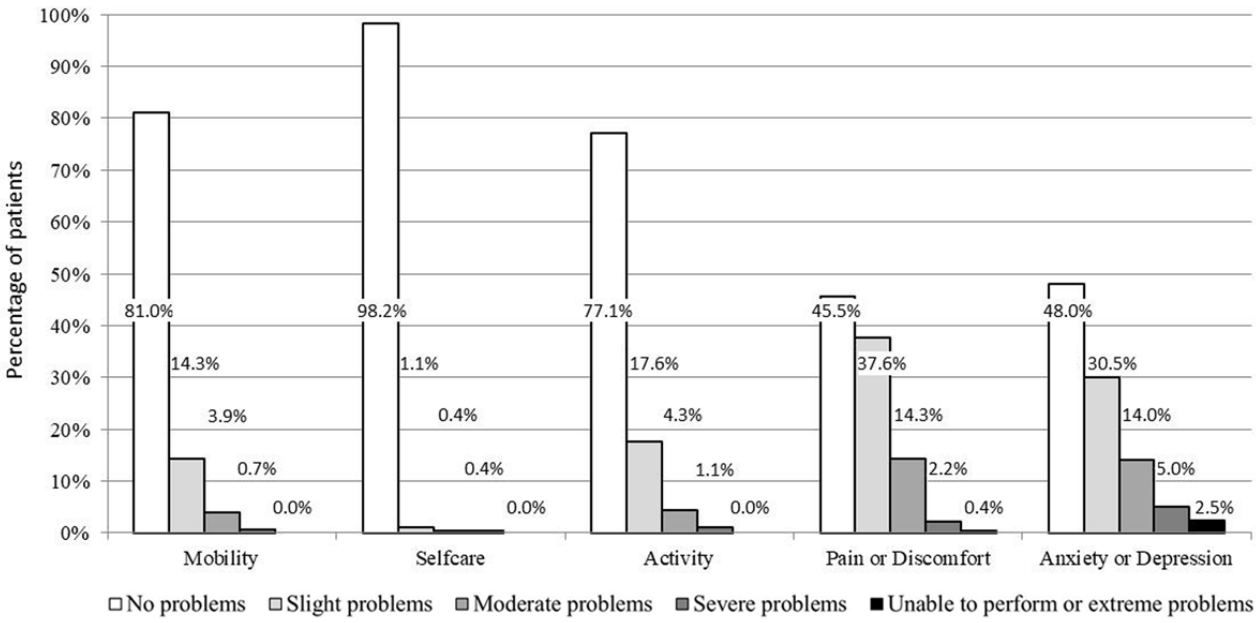


Table 2. The distribution of 5-level EuroQoL-5 dimension version indices (n=279). Cronbach alpha coefficient was .69.

Statistics	EQ-5D-5L[a] (English value set)	EQ-5D-5L (French 3L-5L crosswalk value set)
Theoretical range	–0.28097 to 1	–0.530 to 1
Observed range	0.160 to 1	–0.074 to 1
Mean (SD)	0.88 (0.14)	0.83 (0.19)
Median (interquartile range)	0.92 (0.84 to 1.00)	0.91 (0.71 to 1.00)
Kurtosis (SE)	5.62 (0.29)	3.26 (0.29)
Skewness (SE)	–2.06 (0.15)	–1.63 (0.15)
Floor effect (%)	0	0
Ceiling effect (%)	26.5	26.5

[a]EQ-5D-5L: EuroQol-5 Dimensions-5 Levels.

Table 3. The construct validity of 5-level EuroQoL-5 dimension version.

Constructs	EQ-5D-5L[a] index (English value set)		EQ-5D-5L index (French 3L-5L crosswalk)		EuroQol visual analog scale	
	Mean (SD)	Effect size (95% CI)	Mean (SD)	Effect size (95% CI)	Mean (SD)	Effect size (95% CI)
Other chronic conditions						
0 chronic conditions	0.91 (0.11)	Reference	0.86 (0.14)	Reference	78.91 (14.85)	Reference
1 chronic condition	0.89 (0.10)	0.14 (–0.21 to 0.48)	0.85 (0.15)	0.05 (–0.26 to 0.36)	79.08 (13.23)	0.06 (–0.25 to 0.37)
≥2 chronic conditions	0.82 (0.13)	0.62 (0.18 to 1.06)	0.75 (0.20)	0.60 (0.18 to 1.02)	72.94 (17.22)	0.37 (–0.04 to 0.79)
P value	.002[b,c]	—[d]	.003[b,c]	—	.12	—
Number of SABA[e] canisters prescribed (last year)						
0 canisters	0.89 (0.11)	Reference	0.85 (0.15)	Reference	78.84 (12.90)	Reference
1-4 canisters	0.87 (0.14)	0.11 (–0.18 to 0.39)	0.82 (0.19)	0.19 (–0.07 to 0.45)	76.64 (17.93)	0.21 (–0.06 to 0.47)
5 or more canisters	0.81 (0.17)	0.58 (0.14 to 1.01)	0.76 (0.22)	0.46 (0.05 to 0.86)	72.00 (24.52)	0.47 (0.07 to 0.88)
P value	.02[b]	—	.03[b]	—	.15	—
Frequency of SABA use reported by patient (last 4 weeks)						
Less than once a week	0.82 (0.19)	Reference	0.74 (0.23)	Reference	71.45 (19.85)	Reference
Once or twice a week	0.87 (0.15)	0.17 (–0.11 to 0.44)	0.81 (0.21)	0.29 (0.03 to 0.55)	78.08 (12.92)	0.07 (–0.19 to 0.32)
Almost every day or every day	0.89 (0.12)	0.50 (0.11 to 0.89)	0.85 (0.16)	0.50 (0.15 to 0.84)	78.61 (16.26)	0.37 (0.03 to 0.71)
P value	.03[b]	—	.01[b]	—	.07	—
Asthma Control Questionnaire						
Well-controlled (<0.75)	0.93 (0.10)	Reference	0.91 (0.13)	Reference	81.65 (13.80)	Reference
Intermediate (0.75-1.5)	0.87 (0.11)	0.44 (0.15 to 0.72)	0.81 (0.15)	0.47 (0.22 to 0.73)	79.18 (11.92)	0.15 (–0.11 to 0.40)
Not well-controlled (>1.5)	0.78 (0.19)	1.06 (0.74 to 1.38)	0.69 (0.24)	1.04 (0.75 to 1.32)	68.39 (20.23)	0.79 (0.51 to 1.08)
P value	<.001[b,c,f]	—	<.001[b,c,f]	—	<.001[b,c]	—

[a]EuroQol-5 Dimensions-5 Levels.

[b]First category (reference) versus third category.

[c]Second category versus third category.

[d]*P* value not necessary as the CI was calculated.

[e]SABA: short-acting beta-agonist.

[f]First category (reference) versus second category.

Discussion

Principal Findings

To the best of our knowledge, this is the first study evaluating metric properties of the new EQ-5D-5L in patients with asthma. In this study, this generic preference-based instrument showed adequate distribution and reliability, with 26.5% (74/279) of patients reporting the best possible health state (ceiling effect). In addition, it showed good construct validity, given its capacity of discriminating among groups differing in the number of chronic conditions and symptom control. The distribution of the EQ-5D-5L index was less skewed than the previously published one for the 3-level version owing to its lower ceiling effect [15,17].

Comparison of Web-Based Participation Rate With Prior Work

In this study, 53.7% (312/581) of participants completed the Web-based baseline survey, and almost all of these completed the EQ-5D-5L (92.9%, 290/312). The internet era has led to implementing Web-based surveys to take advantage of the known benefits such as completeness [41,42], low expenses [43], and better data management. Nevertheless, there are still some barriers to Web-based self-completion, which could produce low response rates and selection bias. Although the reported participation rate varied a lot across Web-based surveys [41,44,45], the 53.7% in this study is similar to those reported by other studies comparing different modes of data collection, such as 64.2% and 53.3% participation rates reported by Kongsved et al [41] and Hohwu et al [46] studies. Remarkably, both studies showed a slightly better response rate with the paper mode—73.2% versus 64.2% [41] and 56.2% versus 53.3% [46]. In the ASTRO-LAB cohort, the high overall respondent burden (participants were asked to respond to yearly Web-based surveys, 4-monthly telephone interviews, and monthly short message service text messages) could have affected the response rate.

Comparison With Prior Studies Evaluating the EuroQoL-5 Dimension Version in Patients With Asthma

Despite being higher than the 15% [36] established for the ceiling effect, 26.5% (74/279) of patients with mild-to-moderate persistent asthma in the best possible health state in our sample was considerably lower than that reported in prior studies using the traditional EQ-5D-3L in paper-and-pencil administration [15,17]. A ceiling effect of 59% was described in Japanese patients with mild-to-severe asthma treated with ICs [15], and 50% in Canadian patients with mainly mild-to-moderate self-reported asthma [17]. In addition, our findings showed a lower proportion of patients with no problems in most dimensions than those reported by the 3-level version [15]—81.0% (226/279) versus 90.7% in mobility, 77.1% (215/279) versus 85.2% in activity, 45.5% (127/279) versus 74.1% in pain or discomfort, and 48.0% (134/279) versus 77.8% in anxiety or depression. The other 2 studies on the EQ-5D-3L in asthma [16,17] did not report percentage distributions for each dimension. This lower endorsement of the top response option when compared with results from previous studies with the EQ-5D-3L suggests that the "no problems" category (level 1 out of 3) is partially redistributed to the following intermediate category, "slight" problems (level 2 out of 5), in the new 5-level version. However, head-to-head studies are needed to ensure that the new 5L version's better properties we have observed, compared with results from previous EQ-5D-3L studies [15,17], are not explained by differences in patients' characteristics or design issues.

Studies that directly elicit preferences from representative general population samples to derive value sets for the new EQ-5D-5L, using a harmonized protocol, have already been published for several countries [30,47-51], but they are not yet developed in many others, including France. The EuroQol Group developed the 3L-5L crosswalk value sets as a temporary solution to estimate the EQ-5D-5L in such a situation [29]. The difference between both indices in the negative extreme of the theoretical range (–0.28 and –0.53) is explained by the method used for the elicitation of the societal preference values to derive the value set: time trade-off in the French general population for the 3L version [52], and the composite method of time trade-off with discrete choice experiments in the UK general population for the new 5L version [23,30]. Our findings show that the mean EQ-5D-5L indices obtained with both value sets are quite similar (0.88 and 0.83), supporting that the 3L-5L crosswalk is a good interim solution to calculate the EQ-5D-5L index, until definitive EQ-5D-5L value sets are available.

The EQ-5D-5L index could discriminate among different known groups in the hypothesized direction. In all the variables evaluated, differences between extreme groups ranged from 0.07 to 0.2, therefore being equal to or higher than the minimal important difference, previously estimated as 0.07 [31]. The magnitude was moderate for differences among groups defined by the presence of other chronic conditions and SABA use or prescription and large for differences between patients with well- and not well-controlled asthma measured with the ACQ. McTaggart-Cowan et al [17], with the traditional EQ-5D-3L in patients with asthma, also showed differences between extreme groups >0.07, ranging from 0.07 to 0.18. It was not possible to directly compare effect sizes with this study [17], as the variables to define known groups were different. Mc Taggart-Cowan et al reported a correlation of 0.37 for the ACQ with the EQ-5D-3L index [17], similar to the 0.43 found in our study with the EQ-5D-5L index. These findings indicate a good construct validity for the EQ-5D-5L index, which, in general, presented a greater discriminant capacity than the EQ-VAS among the known groups evaluated.

Limitations and Strengths

Some potential limitations of this study need to be considered. First, a direct comparison with the EQ-5D-3L was not possible. Although previous EQ-5D-3L studies in asthma patients [15-17] showed higher ceiling effects and lower discriminatory properties than ours with the EQ-5D-5L, differences among studies regarding patients' and design characteristics cannot be discarded. Second, as no asthma-specific HRQoL measure was included in this study, we were unable to compare the generic EQ-5D-5L with them. Studies evaluating the EQ-5D-3L in

comparison to the Asthma Quality of Life Questionnaire [15-17] or the Newcastle Asthma Symptoms Questionnaire [16] showed that these disease-specific instruments were more sensitive to change. Further head-to-head studies comparing the EQ-5D-5L with disease-specific instruments are needed, mainly to compare responsiveness. Third, the usability of online versus other methods of survey administration could not be evaluated because all patients completed the Web-based EQ-5D-5L. Fourth, because the ASTRO-LAB project only included patients with mild-to-moderate persistent asthma, the generalizability of our results to those with intermittent or severe persistent asthma is uncertain. The generalizability is also uncertain to patients older than 40 years. Finally, it is important to note that 46.3% (269/581) of participants in the ASTRO-LAB project did not answer the Web-based survey. No differences in sociodemographic characteristics, treatment, and comorbidity were found between respondents and nonrespondents, and differences detected in asthma control were minor. However, there could be differences in other characteristics, which have not been measured such as personality or other psychological traits.

This study has several strengths that need to be highlighted. First, embedding this study in an observational cohort in routine care allowed us to select several appropriate known groups for evaluating the EQ-5D-5L's construct validity in patients with asthma. The relationship between comorbid chronic conditions and health is well established, and the associations of symptoms control [53] with the HRQoL have been extensively studied in this population. Furthermore, the ACQ, validated in 50 languages, is one of the most widely accepted instruments for measuring asthma control [54].

Conclusions

In summary, our results provide support to the construct validity of the EQ-5D-5L administered online to patients with asthma, based on its discriminant ability for distinguishing among health-related known groups, as well as its lower ceiling effect than previously reported for the traditional 3-level version [15,17]. The completion of the EQ-5D-5L by most of the Web-based survey respondents supports the feasibility of this administration form. As it was developed as a preference-based health status measure, the EQ-5D-5L index allows combining both length and quality of life, and calculates QALYs to measure health outcomes in economic evaluations. All these findings suggest that the new EQ-5D with 5 levels is a promising instrument to compare the efficiency of different programs or treatment strategies for asthma patients. Nevertheless, further studies are recommended to evaluate the responsiveness over time of the EQ-5D-5L among patients with asthma.

Acknowledgments

Financial support for this study was provided by the European Union Seventh Framework Programme FP7 (ASTROLAB project EC HEALTH-F5-2011-282593), Instituto de Salud Carlos III FEDER (PI12/00772), Fondo de Investigación Sanitaria, ISCIII (PI15/00449), Fondo de Investigación Sanitaria, ISCIII (PI16/00130), Fondo de Investigación Sanitaria, ISCIII (FI16/00071), ISCIII (Río Hortega, CM15/00167), and DIUE of Generalitat de Catalunya (2017 SGR 452, 2014 SGR 748). The funding agreement ensures the authors' independence in designing the study, interpreting the data, and writing and publishing the report. The authors would like to acknowledge Aurea Martin for her help in the English editing and proofreading process, and in finalizing this manuscript.

Members of the ASTRO-LAB group are: Sandrine Herbage, Manon Belhassen, Marine Ginoux (LBBE, University Claude Bernard Lyon 1, France); Stéphane Schück, Nathalie Texier, Sandy Leproust, Hélène Le Cloarec (Kappa Santé, France); Richard Hubbard (University of Nottingham, England); Alison Bourke, Mary Thompson, Delphine Vial, David Ansell (Cegedim Strategic Data, England); Javier Olaiz, Ana Valcarcel Orti (Lyon Ingénierie Projet, France); and authors MdB, ALD, EVG, LL, MF, OG, and GH.

Authors' Contributions

GH contributed to the conception and design of the paper, conceptualized and oversaw analyses, contributed to the interpretation of data, and wrote the paper. AP contributed to the analysis and gave statistical support. ALD, OG, MMP, JA, EVG, LL, MdB, and KM oversaw all aspects and reviewed the paper for important intellectual content. VSS revised the draft versions of the manuscript. MF oversaw all aspects, contributed to the conception and design of the paper, contributed to the statistical analyses, carried out the interpretation of data, and contributed to the writing of the paper.

Conflicts of Interest

None declared.

References

1. Carranza RJR, Edwards L, Lincourt W, Dorinsky P, ZuWallack RL. The relationship between health-related quality of life, lung function and daily symptoms in patients with persistent asthma. Respir Med 2004 Dec;98(12):1157-1165. [Medline: 15588035]
2. McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. BMC Med 2011 Jul 14;9:86 [FREE Full text] [doi: 10.1186/1741-7015-9-86] [Medline: 21756344]

3. Global Initiative for Asthma. 2018. Global Strategy for Asthma Management and Prevention URL: https://ginasthma.org [accessed 2018-09-26] [WebCite Cache ID 72j3Zhq3B]
4. Stucky BD, Sherbourne CD, Edelen MO, Eberhart NK. Understanding asthma-specific quality of life: moving beyond asthma symptoms and severity. Eur Respir J 2015 Sep;46(3):680-687 [FREE Full text] [doi: 10.1183/09031936.00225014] [Medline: 25882804]
5. Chen H, Gould MK, Blanc PD, Miller DP, Kamath TV, Lee JH, TENOR Study Group. Asthma control, severity, and quality of life: quantifying the effect of uncontrolled disease. J Allergy Clin Immunol 2007 Aug;120(2):396-402. [doi: 10.1016/j.jaci.2007.04.040] [Medline: 17561244]
6. Schatz M, Mosen DM, Kosinski M, Vollmer WM, Magid DJ, O'Connor E, et al. The relationship between asthma-specific quality of life and asthma control. J Asthma 2007 Jun;44(5):391-395. [doi: 10.1080/02770900701364296] [Medline: 17613636]
7. Hunger M, Sabariego C, Stollenwerk B, Cieza A, Leidl R. Validity, reliability and responsiveness of the EQ-5D in German stroke patients undergoing rehabilitation. Qual Life Res 2012 Sep;21(7):1205-1216. [doi: 10.1007/s11136-011-0024-3] [Medline: 21971874]
8. König H, Born A, Günther O, Matschinger H, Heinrich S, Riedel-Heller SG, et al. Validity and responsiveness of the EQ-5D in assessing and valuing health status in patients with anxiety disorders. Health Qual Life Outcomes 2010 May 05;8:47 [FREE Full text] [doi: 10.1186/1477-7525-8-47] [Medline: 20444251]
9. Hung M, Lu W, Chen S, Hou W, Hsieh C, Wang J. Validation of the EQ-5D in Patients with Traumatic Limb Injury. J Occup Rehabil 2015 Jun;25(2):387-393. [doi: 10.1007/s10926-014-9547-0] [Medline: 25261389]
10. EQ-5D. EQ-5D User Guides URL: https://euroqol.org/publications/user-guides [accessed 2018-09-26] [WebCite Cache ID 72j34C3U7]
11. Gibbons E, Fitzpatrick R. Patient Reported Outcome Measurement Group, Oxford. The University of Oxford: Oxford, UK; 2009. A structured review of patient-reported outcome measures (PROMs) for asthma URL: http://phi.uhce.ox.ac.uk/pdf/PROMs_Oxford_Asthma_17092010 [accessed 2018-09-26] [WebCite Cache ID 72j5h8ddn]
12. Räsänen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynänen O, Roine R. Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. Int J Technol Assess Health Care 2006;22(2):235-241. [doi: 10.1017/S0266462306051051] [Medline: 16571199]
13. Richardson J, McKie J, Bariola E. Multi attribute utility instruments and their use. In: Culyer AJ, editor. Encyclopedia of health economics. San Diego: Elsevier Science; 2014:341-357.
14. Fauteux V, Poder T. État des lieux sur les méthodes d'élicitation du QALY Overview of QALY elicitation methods. Int J Health Pref Research 2017;1:001-014. [doi: 10.21965/IJHPR.2017.001]
15. Oga T, Nishimura K, Tsukino M, Sato S, Hajiro T, Mishima M. A comparison of the responsiveness of different generic health status measures in patients with asthma. Qual Life Res 2003 Aug;12(5):555-563. [Medline: 13677500]
16. Garratt AM, Hutchinson A, Russell I. Patient-assessed measures of health outcome in asthma: a comparison of four approaches. Respir Med 2000 Jun;94(6):597-606 [FREE Full text] [doi: 10.1053/rmed.2000.0787] [Medline: 10921766]
17. McTaggart-Cowan HM, Marra CA, Yang Y, Brazier JE, Kopec JA, FitzGerald JM, et al. The validity of generic and condition-specific preference-based instruments: the ability to discriminate asthma control status. Qual Life Res 2008 Apr;17(3):453-462. [doi: 10.1007/s11136-008-9309-6] [Medline: 18274882]
18. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 2004 Sep;13(9):873-884. [doi: 10.1002/hec.866] [Medline: 15362179]
19. Janssen MF, Lubetkin EI, Sekhobo JP, Pickard AS. The use of the EQ-5D preference-based health status measure in adults with Type 2 diabetes mellitus. Diabet Med 2011 Apr;28(4):395-413. [doi: 10.1111/j.1464-5491.2010.03136.x] [Medline: 21392061]
20. van Stel HF, Buskens E. Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease. Health Qual Life Outcomes 2006 Mar 25;4:20 [FREE Full text] [doi: 10.1186/1477-7525-4-20] [Medline: 16563170]
21. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care 2002 Feb;40(2):113-128. [Medline: 11802084]
22. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002 Mar;21(2):271-292. [Medline: 11939242]
23. Oppe M, Devlin NJ, van Hout B, Krabbe PFM, De Leon MC. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health 2014 Jun;17(4):445-453 [FREE Full text] [doi: 10.1016/j.jval.2014.04.002] [Medline: 24969006]
24. Pickard AS, De Leon MC, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Med Care 2007 Mar;45(3):259-263. [doi: 10.1097/01.mlr.0000254515.63841.81] [Medline: 17304084]
25. Kim SH, Kim HJ, Lee S, Jo M. Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea. Qual Life Res 2012 Aug;21(6):1065-1073. [doi: 10.1007/s11136-011-0018-1] [Medline: 21947656]
26. Jia YX, Cui FQ, Li L, Zhang DL, Zhang GM, Wang FZ, et al. Comparison between the EQ-5D-5L and the EQ-5D-3L in patients with hepatitis B. Qual Life Res 2014 Oct;23(8):2355-2363. [doi: 10.1007/s11136-014-0670-3] [Medline: 24627090]

27. Van Ganse E, Texier N, Dima AL, Laforest L, Ferrer M, Hernandez G, ASTRO-LAB group. Assessment of the safety of long-acting β2-agonists in routine asthma care: the ASTRO-LAB protocol. NPJ Prim Care Respir Med 2015 Jun 18;25:15040 [FREE Full text] [doi: 10.1038/npjpcrm.2015.40] [Medline: 26086323]
28. Schatz M, Zeiger RS, Yang ST, Chen W, Crawford WW, Sajjan SG, et al. Relationship of asthma control to asthma exacerbations using surrogate markers within a managed care database. Am J Manag Care 2010 May;16(5):327-333 [FREE Full text] [Medline: 20469953]
29. van Hout B, Janssen MF, Feng Y, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012;15(5):708-715 [FREE Full text] [doi: 10.1016/j.jval.2012.02.008] [Medline: 22867780]
30. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Econ 2018 Jan;27(1):7-22. [doi: 10.1002/hec.3564] [Medline: 28833869]
31. McClure NS, Sayah FA, Xie F, Luo N, Johnson JA. Instrument-Defined Estimates of the Minimally Important Difference for EQ-5D-5L Index Scores. Value Health 2017 Apr;20(4):644-650. [doi: 10.1016/j.jval.2016.11.015] [Medline: 28408007]
32. Reddel HK, Taylor DR, Bateman ED, Boulet L, Boushey HA, Busse WW, American Thoracic Society/European Respiratory Society Task Force on Asthma Control and Exacerbations. An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. Am J Respir Crit Care Med 2009 Jul 1;180(1):59-99. [doi: 10.1164/rccm.200801-060ST] [Medline: 19535666]
33. Taylor DR, Bateman ED, Boulet L, Boushey HA, Busse WW, Casale TB, et al. A new perspective on concepts of asthma severity and control. Eur Respir J 2008 Sep;32(3):545-554 [FREE Full text] [doi: 10.1183/09031936.00155307] [Medline: 18757695]
34. Juniper EF, Svensson K, Mörk A, Ståhl E. Measurement properties and interpretation of three shortened versions of the asthma control questionnaire. Respir Med 2005 May;99(5):553-558 [FREE Full text] [doi: 10.1016/j.rmed.2004.10.008] [Medline: 15823451]
35. Juniper EF, Bousquet J, Abetz L, Bateman ED. Identifying 'well-controlled' and 'not well-controlled' asthma using the Asthma Control Questionnaire. Respir Med 2006 Apr;100(4):616-621 [FREE Full text] [doi: 10.1016/j.rmed.2005.08.012] [Medline: 16226443]
36. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
37. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res 2002 May;11(3):193-205. [Medline: 12074258]
38. Chen W, Lynd LD, FitzGerald JM, Marra CA, Rousseau R, Sadatsafavi M. The added effect of comorbidity on health-related quality of life in patients with asthma. Qual Life Res 2015 Oct;24(10):2507-2517. [doi: 10.1007/s11136-015-0995-6] [Medline: 26038225]
39. Moy ML, Fuhlbrigge AL, Blumenschein K, Chapman RH, Zillich AJ, Kuntz KM, et al. Association between preference-based health-related quality of life and asthma severity. Ann Allergy Asthma Immunol 2004 Mar;92(3):329-334. [doi: 10.1016/S1081-1206(10)61570-0] [Medline: 15049396]
40. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Taylor & Francis, Inc; 1988.
41. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of Internet versus paper-and-pencil versions. J Med Internet Res 2007;9(3):e25 [FREE Full text] [doi: 10.2196/jmir.9.3.e25] [Medline: 17942387]
42. Ritter P, Lorig K, Laurent D, Matthews K. Internet versus mailed questionnaires: a randomized comparison. J Med Internet Res 2004 Sep 15;6(3):e29 [FREE Full text] [doi: 10.2196/jmir.6.3.e29] [Medline: 15471755]
43. Scott A, Jeon S, Joyce CM, Humphreys JS, Kalb G, Witt J, et al. A randomised trial and economic evaluation of the effect of response mode on response rate, response bias, and item non-response in a survey of doctors. BMC Med Res Methodol 2011;11:126 [FREE Full text] [doi: 10.1186/1471-2288-11-126] [Medline: 21888678]
44. Fouladi RT, McCarthy CJ, Moller NP. Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. Assessment 2002 Jun;9(2):204-215. [doi: 10.1177/10791102009002011] [Medline: 12066835]
45. Ekman A, Dickman PW, Klint A, Weiderpass E, Litton J. Feasibility of using web-based questionnaires in large population-based epidemiological studies. Eur J Epidemiol 2006;21(2):103-111. [doi: 10.1007/s10654-005-6030-4] [Medline: 16518678]
46. Hohwü L, Lyshol H, Gissler M, Jonsson SH, Petzold M, Obel C. Web-based versus traditional paper questionnaires: a mixed-mode survey with a Nordic perspective. J Med Internet Res 2013;15(8):e173 [FREE Full text] [doi: 10.2196/jmir.2595] [Medline: 23978658]
47. Augustovski F, Rey-Ares L, Irazola V, Garay OU, Gianneo O, Fernández G, et al. An EQ-5D-5L value set based on Uruguayan population preferences. Qual Life Res 2016 Feb;25(2):323-333. [doi: 10.1007/s11136-015-1086-4] [Medline: 26242249]

48. Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, et al. A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. Med Care 2016 Jan;54(1):98-105 [FREE Full text] [doi: 10.1097/MLR.0000000000000447] [Medline: 26492214]
49. Kim S, Ahn J, Ock M, Shin S, Park J, Luo N, et al. The EQ-5D-5L valuation study in Korea. Qual Life Res 2016 Dec;25(7):1845-1852. [doi: 10.1007/s11136-015-1205-2] [Medline: 26961008]
50. Xie F, Pullenayegum E, Gaebel K, Oppe M, Krabbe PFM. Eliciting preferences to the EQ-5D-5L health states: discrete choice experiment or multiprofile case of best–worst scaling? Eur J Health Econ 2013 Apr 4;15(3):281-288. [doi: 10.1007/s10198-013-0474-3]
51. Versteegh MM, Vermeulen KM, Evers SMAA, de Wit GA, Prenger R, Stolk EA. Dutch Tariff for the Five-Level Version of EQ-5D. Value Health 2016 Dec;19(4):343-352 [FREE Full text] [doi: 10.1016/j.jval.2016.01.003] [Medline: 27325326]
52. Chevalier J, de Pouvourville G. Valuing EQ-5D using time trade-off in France. Eur J Health Econ 2013 Feb;14(1):57-66. [doi: 10.1007/s10198-011-0351-x] [Medline: 21935715]
53. King MT, Kenny PM, Marks GB. Measures of asthma control and quality of life: longitudinal data provide practical insights into their relative usefulness in different research contexts. Qual Life Res 2009 Apr;18(3):301-312. [doi: 10.1007/s11136-009-9448-4] [Medline: 19225906]
54. Worth A, Hammersley V, Knibb R, Flokstra-de-Blok B, DunnGalvin A, Walker S, et al. Patient-reported outcome measures for asthma: a systematic review. NPJ Prim Care Respir Med 2014 Jun 26;24:14020 [FREE Full text] [doi: 10.1038/npjpcrm.2014.20] [Medline: 24964767]

Abbreviations

ACQ: Asthma Control Questionnaire
EQ-5D-5L: 5-level EuroQoL-5 dimension version
EQ-VAS: EuroQol-visual analog scale
HRQoL: Health-Related Quality of Life
IC: inhaled corticosteroid
LABA: long-acting beta-agonists
SABA: short-acting beta-agonists

Edited by G Eysenbach; submitted 14.03.18; peer-reviewed by T Poder, D Williams, M Duplaga; comments to author 03.08.18; revised version received 27.09.18; accepted 03.10.18; published 23.01.19.

Please cite as:
Hernandez G, Garin O, Dima AL, Pont A, Martí Pastor M, Alonso J, Van Ganse E, Laforest L, de Bruin M, Mayoral K, Serra-Sutton V, Ferrer M, ASTRO-LAB Group
EuroQol (EQ-5D-5L) Validity in Assessing the Quality of Life in Adults With Asthma: Cross-Sectional Study
J Med Internet Res 2019;21(1):e10178
URL: http://www.jmir.org/2019/1/e10178/
doi:10.2196/10178
PMID:30672744

Original Paper

A Modular Health-Related Quality of Life Instrument for Electronic Assessment and Treatment Monitoring: Web-Based Development and Psychometric Validation of Core Thrive Items

Paul Wicks[1], PhD; Stacey McCaffrey[1], PhD; Kim Goodwin[1], BA; Ryan Black[1], PhD; Michael Hoole[1], BS; James Heywood[1], SB

PatientsLikeMe, Cambridge, MA, United States

Corresponding Author:
Paul Wicks, PhD
PatientsLikeMe
160 2nd Street
Cambridge, MA, 02142
United States
Phone: 1 6174994003
Email: pwicks@patientslikeme.com

Abstract

Background: Patient-reported outcome (PRO) measures describe natural history, manage disease, and measure the effects of interventions in trials. Patients themselves increasingly use Web-based PRO tools to track their progress, share their data, and even self-experiment. However, existing PROs have limitations such as being: designed for paper (not screens), long and burdensome, negatively framed, under onerous licensing restrictions, either too generic or too specific.

Objective: This study aimed to develop and validate the core items of a modular, patient-centric, PRO system (*Thrive*) that could measure health status across a range of chronic conditions with minimal burden.

Methods: Thrive was developed in 4 phases, largely consistent with Food and Drug Administration guidance regarding PRO development. First, preliminary core items (common across multiple conditions: *core Thrive items*) were developed through literature review, analysis of approximately 20 existing PROs on PatientsLikeMe, and feedback from psychometric and content experts. Second, 2 rounds of cognitive interviews were iteratively conducted with patients (N=14) to obtain feedback on the preliminary items. Third, core Thrive items were administered electronically along with comparator measures, including 20-item Short-Form General Health Survey (SF)-20 and Patient Health Questionnaire (PHQ)-9, to a large sample (N=2002) of adults with chronic diseases through the PatientsLikeMe platform. On the basis of theoretical and empirical rationale, items were revised or removed. Fourth, the revised core Thrive items were administered to another sample of patients (N=704) with generic and condition-specific comparator measures. A psychometric evaluation, which included both modern and classical test theory approaches, was conducted on these items, and several more items were removed.

Results: Cognitive interviews helped to remove confusing or redundant items. Empirical testing of subscales revealed good internal consistency (Cronbach alpha=.712-.879), test-retest reliability (absolute intraclass correlations=.749-.912), and convergent validity with legacy PRO scales (eg, Pearson r=.5-.75 between Thrive subscales and PHQ-9 total). The finalized instrument consists of a 19-item core including 5 multi-item subscales: *Core symptoms*, *Abilities*, *Mobility*, *Sleep*, and *Thriving*. Results provide evidence of construct (content, convergent) validity, high levels of test-retest and internal consistency reliability, and the ability to detect change over time. The items did not exhibit bias based on gender or age, and the items generally functioned similarly across conditions. These results support the use of Thrive Core items across diverse chronic patient populations.

Conclusions: Thrive appears to be a useful approach for capturing important domains for patients with chronic conditions. This *core* set serves as a foundation to begin developing modular condition-specific versions in the near future. Cross-walking against traditional PROs from the PatientsLikeMe platform is underway, in addition to clinical validation and comparison with biomarkers. Thrive is licensed under Creative Commons Attribution ShareAlike 4.0.

(J Med Internet Res 2019;21(1):e12075) doi:10.2196/12075

KEYWORDS

personal health records; health-related quality of life; patient reported outcome measures

Introduction

Patient-Reported Outcomes

Patient-reported outcomes (PROs) are reports of health status that come directly from the patient and are typically captured via a questionnaire that has been developed with clearly defined methods, provides proof of validation, and has instructions for use [1]. PROs are one method of incorporating patient perspectives into drug development [2], such as helping to identify trade-offs between treatment characteristics and health-related quality of life (HRQoL) [3]. Accordingly, academic researchers, clinicians, pharmaceutical manufacturers, and their contract research organizations have developed over a thousand PROs over the past few decades with the intent to use some of them as endpoints within clinical trials [4,5]. PROs include single-domain and multi-domain instruments covering a diverse array of domains including overall health status, condition impact, HRQoL, mood, pain, functioning, medication adherence, and treatment side effects.

In addition to their use in trials, a subset of (mostly specialist) clinics deploy PROs during routine clinical practice to help monitor patient symptoms and functioning and to assist with decision making. The incorporation of PROs into electronic medical records is likely to accelerate this trend [6]; their use for symptom management has been particularly successful in oncology [7]. Routine use of remote symptom monitoring is associated with clinically significant benefits in HRQoL, fewer admissions, and even overall survival, probably via improved communication with health care professionals [8].

Whereas other medical tools such as continuous glucose monitors were once the preserve of specialist clinics to check on patient compliance, today people with diabetes themselves are using these tools and integrating them into self-coded apps and jury-rigged mechanisms to develop their own *closed-loop open artificial pancreas* [9]. It should be no surprise, then, that some patients and caregivers harness PROs, research tools originally designed to monitor the outcomes of whole groups of patients in clinical trials, and use them to understand their own individual progress with disease, put themselves into context, self-experiment, and even conduct citizen-science experiments [10]. With the right support, some patients have even developed their own PROs to deal with the frustrations they have encountered with repurposing tools to suit their needs [11].

That was part of the motivation behind the development of the online community PatientsLikeMe, which was first founded in 2005. One feature of the site allows people living with amyotrophic lateral sclerosis (ALS) to access a patient-reported version of the clinician-reported outcome (ClinRO) used in clinical research to characterize patient function, the ALS functional rating scale revised (ALSFRS-R [12]). At the time PatientsLikeMe was launched, ALS researchers were advised not to tell research participants their own ALSFRS-R scores or how they were doing relative to other patients like them [13]. Patients tracking their own ALSFRS-R scores on the site could see their progression overlaid on percentile curves of other patients like them (with different curves for slower ALS subtypes such as progressive muscular atrophy and primary lateral sclerosis) and bring these data to clinic appointments with their health care professionals, helping to improve communication and management [14]. At first, there was concern that PROs might lack resolution and accuracy relative to ClinROs, yet subsequent validation studies have found a high degree of agreement (eg, Spearman rho=.965, $P \leq .001$ [15]).

Limitations of Patient-Reported Outcomes for Digital Health Apps

However, as PatientsLikeMe expanded to other conditions such as multiple sclerosis (MS), Parkinson disease (PD), HIV, mood disorders, fibromyalgia, epilepsy, autism spectrum disorder, and organ transplants, it became clear that the state of PRO development was highly uneven across these conditions. While some PROs focused on symptoms and pathological elements of disease, others focused on the impact of the condition, treatment side effects, or broader concepts such as HRQoL. As standards on the quality of PRO development (such as the Food and Drug Administration's (FDA's) guidance for industry on PRO development in labeling [16]) became available, it also became clear that the psychometric quality and rigor of instruments varied enormously, with some meeting only low standards of reliability, having had little input from patients themselves, or undergoing little in the way of psychometric validation for responsiveness to change, clinically important differences, or conformity to modern psychometric methods such as Rasch modeling [17]. In addition to well-worn limitations identified in the psychometric field [17], we identified a range of issues that may not have arisen in traditional clinical settings but are problematic for their use in digital health apps for both patients (Table 1) and professionals (Table 2).

Objectives of This Study

Adapting what we felt were the best approaches from the PRO field, we sought to develop a *modular* questionnaire system that addressed the limitations we had identified for their use in real-world and digital health apps. Specifically, we aimed to develop a set of questions that covered the key domains of HRQoL in adults with chronic illness that was brief, minimally burdensome, positively framed, and that could interleave additional items to account for comorbidity in future condition-specific modules.

Methodologically, we sought to conform (to the extent possible) with the FDA's Guidance for Industry for PRO development [16] by completing the following objectives:

1. Developing a conceptual framework and the preliminary item pool through literature review and expert input
2. Cognitive debriefing of draft items with participants
3. Revising these items and framework accordingly
4. Collecting data and evaluating psychometric properties (such as rating scale functioning, reliability, convergent validity, ability to detect change, and bias)
5. Modifying the instrument based on results of the empirical evaluation
6. Collecting data and analyzing psychometric properties of the revised instrument
7. Finalizing the instrument and scoring

Table 1. Issues identified by the team for patients with the patient-reported outcome status quo.

Issues for patients	Example in existing PROs[a]	Implications	Proposed solution	Implementation in Thrive
PROs ignore comorbidity	For example, SF-36[b] does not contain important domains for a specific chronic condition, whereas condition-specific instruments are unclear on how user should dissociate primary condition from comorbidities	Typical PatientsLikeMe user has a median of 3 moderate-serious medical conditions; fielding additional PROs for each condition dramatically increases burden and redundancy	Core Thrive items asked of all users; curated set of additional symptoms, abilities, and thriving items fielded according to reported conditions	Core Thrive item asks separately about impact of each condition and comorbidity independently, for example, "Parkinson's impact=a lot" but "Eczema=not at all"
No personalization for the individual	Redundant questions, for example, pregnancy in males. At best, there are instructions to skip irrelevant questions (eg, "If no, skip to 12")	Patients wade through the same clumsy skip logic instructions (or irrelevant questions) over and over again	Let patients specify once that something is not relevant and remember that in the future	Option of "Stop asking me this" checks why patient wants to skip and asks if we can assume the last answer given will continue being the same
Large number of questions	For example, autism treatment evaluation checklist contains 78 items	Takes a long time to complete (approximately 10 seconds per item) and may cause drop-off	Ask as few questions as possible	Review of literature and patient-submitted data to identify most common issues
Long question stems and responses	Parkinson disease rating scale requires reading 1456 words	Difficult to read on mobile screens, may require scrolling, risks biasing answers	Use brief, active voice items and consistent response scales rather than longer text-anchored responses	Items are Likert-style unipolar responses
Negative framing	For example, Beck Depression Inventory: "(0) I don't feel disappointed in myself (1) I am disappointed in myself (2) I am disgusted with myself (3) I hate myself"	Fails to identify, for example, users who feel good about themselves; ignores islands of resilience and important self-expression for users; not appealing to use repeatedly	Frame items in a positive or at least neutral way when possible	Abilities stem asks, "how well could you" and Thriving stem asks, "how often could you"
Variable or unclear recall periods	Recall periods may be missing, "past week" vs "past 7 days", or very long, for example, past 12 months or "since you were diagnosed"	Different user needs require different recall periods	Codify and test different response periods flexibly, that is, "In the past <recall period> how well could you <activity>?"	Initial validation study developed with "last month" recall period but future work will test other recall periods
Potentially sexist items	For example, fibromyalgia impact questionnaire focuses on disease preventing patient from doing shopping, laundry, and housework	Risks offending users. Also ignores modern options such as home grocery delivery	Avoid making assumptions about how people live their lives with or without illness	Provide general role function items, for example, "responsibilities" or "personal needs" rather than specific chores
Anachronistic items	For example, adolescent systemizing spectrum quotient asks about "programming a video recorder"	Unclear how users will interpret such items; potential for user frustration	Focus on personally defined impact of condition rather than task completion	Use *evergreen* items such as walking or sleeping
Confusing scores and directionality across conditions PROs	For example, scores such as the ALSFRS-R have an arbitrary range 0-48, Unified Parkinson's Disease Rating Scale is 0-199; sometimes higher is worse, sometimes lower	Difficult for patients to understand meaning; conveys false sense of an interval or ratio level scale	Use a score based on a more relatable frame of reference, for example, 0-10	10-point scales are more familiar

[a]PRO: patient-reported outcome.

[b]SF-36: short-form 36 questionnaire.

Table 2. Issues identified by the team for professionals with the patient-reported outcome status quo.

Issues for professionals	Example in existing PROs[a]	Implications	Proposed solution	Implementation in Thrive
Incomplete documentation	Most instruments lack detailed instructions for missing data	Unclear how to score, where more validation work is needed, whether items contain bias	Digitize and share item-level response characteristics through data repositories	Work in progress
Onerous licensing restrictions	For example, license-holders of *Morisky medication adherence scale* have threatened lawsuits, demanded fees, and required retractions for an 8-item questionnaire	Risk of litigation restricts innovation. Digital health practitioners may need to adapt licensed instruments to their own needs without wanting to revalidate entire instrument.	All PROs should be licensed under *Creative Commons ShareAlike* to promote scientific dissemination and innovation so that anyone can use and modify them, for free, forever	All Thrive items and supporting documentation are licensed under Creative Commons ShareAlike 4.0

[a]PRO: patient-reported outcome.

Methods

Each phase of the instrument development and validation study is presented in temporal sequence below (Figure 1).

Setting

Participants were recruited from the membership of PatientsLikeMe.com, an online community for patients living with chronic illness. Potential members are made aware of the site through a variety of channels including Web-based advertising, nonprofit partners, word of mouth, and search. Members join the site with a goal to find other patients like them, track their condition over time, and to benefit from the shared experiences of other members like them [18]. The site is currently only available in English, with most participants living in the United States. Participants were not offered any reimbursement for participating in this study. As a convenience sample of chronic online patients, this group is representative of digital health patients, but caution should be taken in generalizing these findings to other groups.

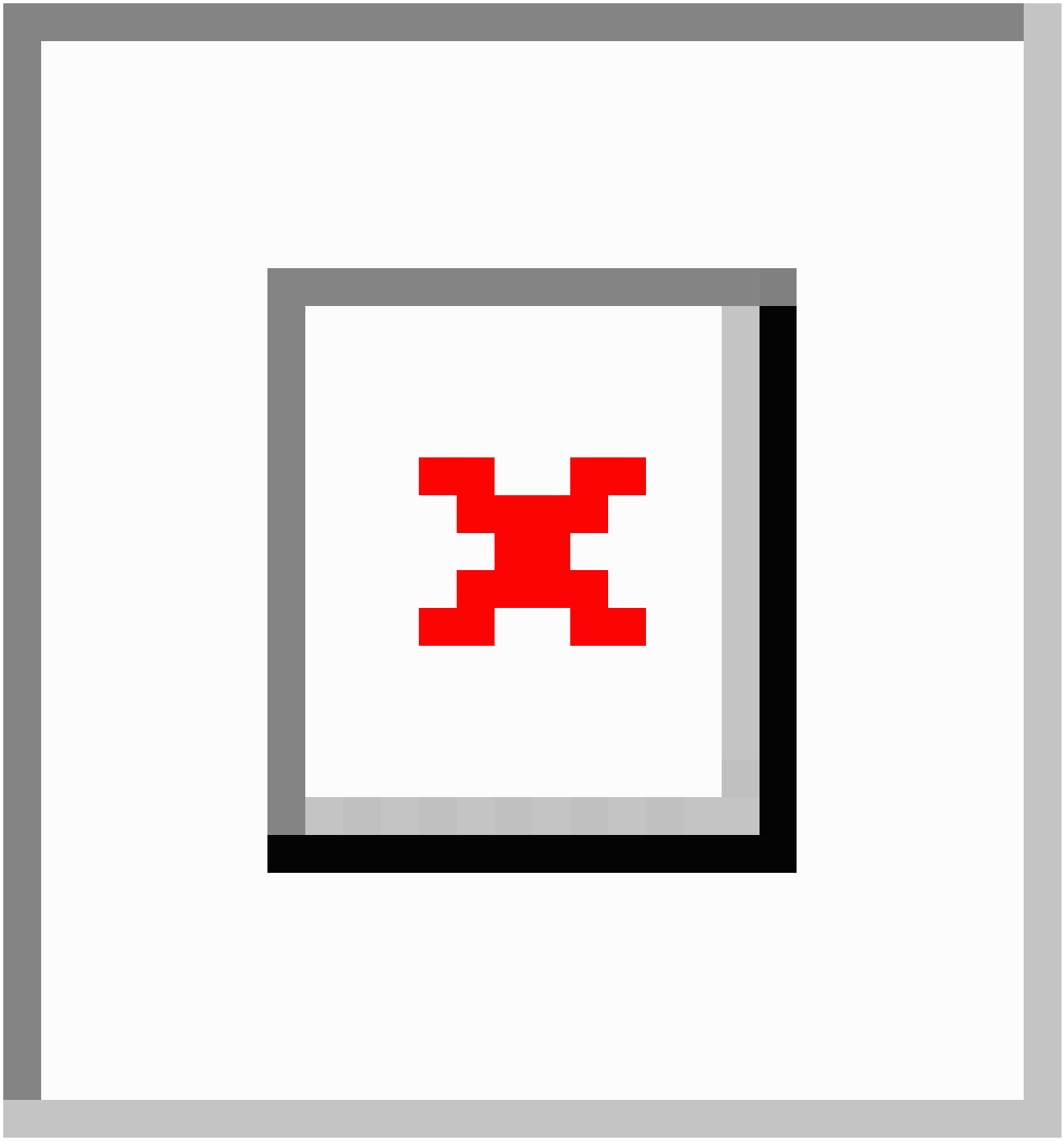
Ethical Approval

On request for ethical independent review board, this research was exempted from further ethical review by the New England Independent Review Board as a minimal risk study (WO 1-2559-1).

Developing a Conceptual Framework

A literature search was conducted to guide the development of a preliminary conceptual model and item generation. Consistent with widely regarded conceptual models [19,20], HRQoL was considered to be a broad and dynamic construct that incorporates quality of life, general health perceptions, functional status, symptoms, as well as intraindividual and environmental factors.

Figure 1. Overview of validation process, adapted from the Food and Drug Administration (2009) guidance for industry. PLM: PatientsLikeMe.



Each of these aspects was considered when developing the initial item pool to ensure that the final Thrive core items adequately captured HRQoL. In particular, we were influenced by the Patient-Reported Outcomes Measurement Information System (PROMIS) Domain Framework [21] and prospectively sought to develop items relevant to physical health (including common symptoms such as pain, fatigue, and sleep disturbance), mental health (including mood symptoms, cognitive dysfunction, and positive psychology), and social health (including ability to participate and social isolation). The research team, which included content experts and psychometricians, collaboratively drafted the preliminary Thrive items, some of which were adapted from validated instruments and published PRO HRQoL tools (eg, the *SF-1* general health item from the Research and Development (RAND) Corporation SF-36 [22]).

Item Development

A PRO instrument consists of instructions, items (which incorporate a recall period), and the items' response options. Given the focus on chronic health conditions, we settled on a *last month* response window. Although a *30-day* response window may appear more precise, we aimed for questions to seem conversational. Since we planned to code questions as medical objects in a database to support use across multiple platforms, wherever possible, we tried to take a consistent approach to question stems and response options.

Symptoms were defined as any physical or mental feature regarded as indicating a condition or disease, particularly when such a feature was apparent or bothersome to the patient. On the basis of the World Health Organization's International Classification of Function [23], we offered participants a consistent symptom question type: "Please rate the severity of any <SYMPTOM ITEM> over the past month" and response options: *None*, *Mild*, *Moderate*, or *Severe*.

Abilities were defined as the degree to which a participant possessed the means to do something important to them, particularly to function independently. On the basis of our aim to offer positively framed question stems, we phrased these as "Over the last month, how well could you <ABILITY ITEM>?" with response options of *Extremely well*, *Very well*, *Fairly well*, *Poorly*, or *Not at all*. In this way, we aimed to identify participants who were functioning particularly well on some items despite their condition as well as to make the experience of taking the instrument a more pleasant one, and to avoid floor or ceiling effects.

Thriving was defined as the extent to which a participant was living the life they wanted to lead, regardless of their health status. These were phrased "Over the last month, how often did you <THRIVING ITEM>?" with a response scale of *All of the time*, *Most of the time*, *Some of the time*, or *None of the time*.

Cognitive Interviews

Procedures

Cognitive interviews were conducted to gather qualitative feedback regarding the preliminary items and to establish content validity. A total of 2 interviewers trained in cognitive interviewing procedures completed the interviews individually with participants over the phone. Interviews were not audio-recorded and lasted approximately 90 min. Retrospective probing was used to enhance realism [24], and interviewers followed a semistructured interviewing script that allowed for deviation as appropriate. Cognitive interviews were conducted in 2 rounds so that content modified following the first round of interviews could be evaluated in a second round.

Participants

As one of the main objectives was to create a system that would replace the legacy PROs on the PatientsLikeMe website, to ensure that the items were reviewed by a diverse patient group living with chronic health conditions who were representative of our most populated communities, members of PatientsLikeMe who met the following study inclusion criteria were invited to participate:

1. Reported a primary condition of ALS, PD, multiple sclerosis (MS), major depressive disorder (MDD), generalized anxiety disorder (GAD), or posttraumatic stress disorder (PTSD)
2. Aged 18 years or older
3. Primarily resided in the United States

Empirical Evaluation

Following cognitive interviews, the draft core Thrive items were programmed in PatientsLikeMe's research survey tool (RST) and administered along with validated comparison measures (PHQ-9 and the Medical Outcomes Study SF-20) to patients with chronic medical conditions (*Round 1*). On the basis of the items' psychometric functioning and expert input, items were revised or removed. The updated Thrive instrument was again administered to an independent sample of patients (*Round 2*) alongside validated generic comparison measures (PHQ-9, SF-20) and PROs offered to patients on the PatientsLikeMe website with at least some psychometric validation (multiple sclerosis rating scale, MSRS) for participants with MS, ALSFRS-R for participants with ALS, and PatientsLikeMe-QoL for participants with systemic lupus erythematosus (SLE). Additional PROs used on PatientsLikeMe were fielded (Parkinson's disease rating scale [PDRS] in PD and *mood map* in mood disorders) but because of a lack of previous psychometric validation, they are not reported here.

During both rounds, participants were asked to complete assessments at 3 timepoints:

1. (Administration 1) Thrive + comparator measures, baseline
2. (Administration 2) Thrive only: 3 days after Administration 1, for evaluating stability
3. (Administration 3) Thrive + comparator measures: 30 days after Administration 1, for evaluating ability to detect change over time

Materials-Comparator Patient-Reported Outcomes

Patient Health Questionnaire-9: All Participants

The PHQ-9 is a 9-item self-report measure of depression based on the Diagnostic and Statistical Manual, Fourth Edition diagnostic criteria [25]. It has been validated for use with primary care, obstetrics/gynecological patients, and the general population, and has been found to be useful as both a clinical and research tool [25-27]. It has also demonstrated sensitivity to detect change in depression status over time in medical outpatients [28].

Short-Form General Health Survey -20: All Participants

SF-20 is a brief self-report health survey that captures 6 health concepts: physical functioning, role functioning, social functioning, health perceptions, pain, and mental health [29]. The SF-20 has exhibited adequate levels of reliability and validity in a general population sample and patient population [29,30].

Multiple Sclerosis Rating Scale: Participants With Multiple Sclerosis

Inspired by the Guy's Neurological Disability Scale [31], the 7-item MSRS was developed by PatientsLikeMe to capture the impact of MS on daily living. This scale has demonstrated convergent validity through correlations with walking scores and physician-derived measures [32].

Amyotrophic Lateral Sclerosis Functional Rating Scale Revised: Participants With Amyotrophic Lateral Sclerosis

The ALSFRS-R is one of the most widely used instruments to capture ALS disease progression [12]. The ALSFRS-R is correlated with disease progression and survival [33,34], and

research has suggested good internal consistency and reproducibility.

PatientsLikeMe-Quality of Life: Participants With Systemic Lupus Erythematosus

The PatientsLikeMe-QoL is intended to capture HRQoL related to physical function, mental distress, and social functioning over the past 30 days. This instrument has exhibited high internal consistency and convergent validity [35].

Power Analysis

The target N for each patient group at administration 2 (3-day retest) was 100. The sample size of 100 was derived from a power analysis to detect a significant difference between an intraclass correlation coefficient of .80 (within the acceptable level) and 0.69 (below the acceptable), assuming 80% power. Specifically, a sample size of 100 would detect whether the CI of the reliability coefficient includes values below the accepted reliability threshold (Rxx=.70) 80% of the time. Notably, because of difficulties with achieving a sufficient sample size during Round 2, results were not evaluated separately by patient group.

Participants

Adult (18 years or older) PatientsLikeMe members primarily residing in the United States who reported a primary condition of ALS, MS, PD, MDD, GAD, PTSD, or SLE were sent an invitation to participate through the PatientsLikeMe platform. The following information is reported in accordance with the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) checklist [36]. All surveys were voluntary and would not affect invitees' use of other features on the PatientsLikeMe site. Individual users had a password-protected log-in and could only take the survey once; we have tools to prevent multiple accounts from originating in the same location, including account registration, cookies, and internet provider tracing. No incentives were offered, question order was not randomized, certain items only appeared based on responses to previous questions (ie, were *branching*) to minimize burden, and the total number of questions varied per respondent. There was 1 question per page with a *back* button allowing patients to navigate back 1 page to review their previous response.

On March 23, 2017, 20,941 PatientsLikeMe members fitting the inclusion criteria mentioned above were invited to the Round 1, baseline survey; this survey remained open until April 10, 2017. Participants who did not complete this survey were sent 1 reminder message 3 days after the invitation. Those who completed the survey were automatically sent an invitation to administration 2 three days after completion of administration 1. Administration 2 was open for the same time period as the baseline survey. Those who completed the Round 1 baseline survey were invited to a 30-day retest (administration 3) on May 2, 2017, which remained open until May 10, 2017.

For the second round of the surveys, 12,460 participants were sent invitations on June 15, 2017, to the Round 2 baseline survey, which remained open until July 6, 2017. Reminders and the 3-day test/retest invitation were sent in a manner identical to that of Round 1; Round 2-administration 2 was also open from June 15, 2017, to July 6, 2017. Those who completed the Round 2 baseline were invited to a 30-day retest on July 25, 2017 which remained open until August 10, 2017. All numbers pertaining to Round 1 and Round 2 are reported in the results section.

Analytic Plan

Psychometric validation is an iterative process that is driven by both theoretical and empirical support; therefore, the Thrive research team provided input and feedback during each step of the validation process. Thrive was evaluated using both classical and modern test theory approaches, including evaluation of: rating scale functioning, dimensionality, person-to-item targeting, bias (gender [male, female], race [white, nonwhite], condition [neurodegenerative, autoimmune relapsing, psychiatric]), internal consistency reliability, test-retest reliability, convergent validity, and ability to detect change using longitudinal data. The primary purpose of the first round of testing was to explore item functioning and to make revisions as necessary before the second round. Analytic procedures for this second round of testing were largely consistent with those utilized in Round 1. Readers are referred to Bond and Fox [37] and Furr [38] for more information about these analyses. Analyses were conducted in SPSS version 24 (IBM Corporation, New York) and WINSTEPS 3.74.0 (Beaverton, Oregon) by author SM.

Results

Cognitive Interviews

Twelve participants completed the first round of cognitive interviews. Participants (75% [9/12] female) reported primary diagnoses of MS (33% [4/12]), fibromyalgia (17% [2/12]), GAD (8% [1/12]), MDD (8% [1/12]), ALS (8% [1/12]), bipolar disorder (8% [1/12]), and SLE (8% [1/12]). As cognitive interviews were being conducted, the interviewers regularly met together and with the research team to discuss participant feedback with the goal of identifying recurring themes. Participants identified several items that had redundant content, were too vague and caused confusion, or that they felt were not important for purposes of monitoring their health. Several items were removed or revised based on participants' suggested rewordings to increase clarity, response options were modified to enhance consistency or reduce confusion, and the recall period was made consistent across items. For example, when probed about a coping question ("How well could you cope over the last month?"), participants expressed confusion (eg, "Cope with what?") and felt that one's ability to cope and deal with life stressors was already covered by other items. Similarly, response options of several items were modified for consistency and to reduce confusion. For example, the question wording "How well could you see yourself as a worthwhile person over the last month?" was changed to "Over the last month, how often did you see yourself as a worthwhile person?"

A few respondents wanted to express more detail about pain or sleep, which were issues of particular concern for them. As this core instrument is meant to be applicable to all PatientsLikeMe members, the research team decided to revisit further detail on those issues as future modular additions to the instrument.

A second round of cognitive interviewing was conducted to evaluate the revised content. A total of 2 participants (1 male) completed the second round of cognitive interviews. These participants reported primary diagnoses of bipolar disorder and SLE. Participants provided relatively similar and positive feedback about the items. This feedback was communicated back to the research team and minor revisions to the survey were made.

Psychometric Evaluation: Round 1

Participants

Consolidated Standards of Reporting Trials (CONSORT) flow diagrams are presented in Multimedia Appendix 1. Of the 20,941 PatientsLikeMe members who were sent an invitation to participate, 2311 responded to the invitation by clicking on the survey link, and 86.6% of these members completed administration 1 (N=2002). It took participants approximately 16 min to complete the battery. The survey was open for 17 days in total. To evaluate test-retest reliability, the same 2002 participants were invited to complete the draft core Thrive items 3 days after the first administration. The retest was completed by 924 participants. Finally, the original 2002 participants were invited to complete the battery (Thrive, PHQ-9, SF-20) approximately 30 days after the first and initial administration to evaluate core Thrive items' ability to detect change over time. In total, 717 participants completed the battery at the 30-day administration.

Demographic and clinical characteristics of this sample are presented in Table 3. The average age of participants was approximately 55 years, and the majority of participants were non-Hispanic, white, and female. The sample was highly educated; 33% of the sample completed some college, 25% completed college, and 19% received postgraduate education.

Round 1 Results

The purpose of Round 1 was to explore item functioning and to make revisions as necessary before the second round. A summary of results from Round 1 can be found in Multimedia Appendix 1. Some of the scales evidenced levels of reliability that are below what is typically considered acceptable, some items exhibited bias or poor discrimination. Core Thrive items were modified based on these findings and were subject to empirical evaluation in Round 2.

Psychometric Evaluation: Round 2

Participants

Of the 12,460 participants who were sent an invitation to participate, 887 responded to the invitation by clicking on the survey link, and 79.4% of these participants (N=704) completed the Round 2 baseline survey; 239 completed the 3-day retest and 51 completed 30-day retest. Demographic and clinical characteristics of this sample are presented in Table 4.

Round 2 Results

Results are presented by scale below and are summarized in Tables 5-7. The final surviving Thrive items from Round 2 testing are listed in Table 8, and summary of the items that were retained or discarded is provided in Multimedia Appendix 2. Detailed results, including evaluation of dimensionality, item difficulty, fit statistics, response category thresholds, and person-to-item maps, are also presented in Multimedia Appendix 2.

Empirical testing of subscales revealed good internal consistency (Cronbach alpha=.712-.879) and test-retest reliability (absolute intra class correlations=.749-.912). Cronbach alpha for the Sleep subscale was lower (Cronbach alpha=.712), probably owing to the lower count of items.

Convergent validity varied by domain. Correlations were highest between the *Overall Health* Thrive item and *General Health Item* of the SF-20 owing to the similarity of their stem phrasing (Thrive: "Over the last month, how has your health been?", SF-20: "In general, would you say your health is?") with the same response options but with different response time periods. The *Impact of Primary Condition* had consistent moderate correlations with all comparator measures (Pearson r=.443-.518). Core symptoms (including anxious mood, depressed mood, fatigue, pain, and stress) had stronger correlations with mental health comparators (Pearson r=.-750-.775 for PHQ-9, SF-20 mental health, PLM-QoL mental subscale) than physical health comparators (Pearson r=.390-.698 for SF-20 physical function, PLM-QoL physical, nonsignificant with ALSFRS-R). The single-item Mobility scale (Walking) had a moderate correlation with physical functioning comparators that themselves contained walking items (SF-20 physical function, MSRS, PLM-QoL physical scale, ALSFRS-R). The Abilities scale correlated most strongly with the PLM-QoL (Pearson r=.770-.809), which asks participants to endorse the extent to which their health limited their ability to participate in physical functioning, mental well-being, or social interaction. Two psychological items (Cognitive and Emotional control) may explain the relatively high correlation with the PHQ-9 (Pearson r=.744). Abilities had a moderate degree of correlation (Pearson r=.450-.520) with comparator measures of physical role function or physical ability (SF-20 physical role, ALSFRS-R). Thriving items were most strongly related to mental health comparators (Pearson r=.743-.806 for PHQ-9, SF-20 mental health, PLM-QoL mental) but had nonsignificant or weak correlations with physical health comparators (Pearson r=.342 for SF-20 physical health, r=.132 *P*=.32 with ALSFRS-R).

Analysis of longitudinal residualized change scores over 30 days found significant, but attenuated, patterns of correlation similar to the results of the convergent validity analysis. The strongest relationship (Pearson r=.496) was between the 2 item-Sleep scale (Falling asleep and Staying asleep) with the single-item PHQ-9 question.

Overall Health

Absolute agreement of responses across the 3-day test-retest period (n=239) suggested adequate stability (Table 5). Convergent validity was evaluated by calculating a Pearson correlation between Overall Health and the SF-20 General Health item. Results yielded a strong correlation, providing support for the convergent validity of the Overall Health scale (Table 6). Next, ability to detect change was evaluated by correlating residualized change scores of Overall Health and the SF General Health item over the 30-day testing period.

Stated differently, we evaluated the correspondence between change in patients' responses over time. Results supported the Overall Health scale's ability to detect change over time (Table 7).

Impact of Primary Condition

Absolute agreement of responses to the Impact of Primary Condition item across the 3-day test-retest period was adequate (Table 5). The Impact of Primary Condition scale was related as anticipated to comparator measures, providing support for convergent validity (Table 6). Correlations between residualized change scores (see Table 7) provide support for the Impact of Primary Condition scale's ability to detect change over time.

Table 3. Round 1 participant demographics.

Variable	Baseline	3-day test-retest	30-day retest
Participants (n)	2002	924	717
Age (years), mean (SD)	54.9 (11.6)	56.2 (10.7)	56.0 (11.3)
Conditions, median (range)	2 (1-58)	2 (1-53)	2 (1-58)
Gender, n (%)[a]			
Male	600 (30.0)	290 (31.5)	245 (34.2)
Female	1399 (70.0)	632 (68.5)	471 (65.8)
Ethnicity, n (%)[a]			
Hispanic	77 (4.0)	31 (3.5)	26 (3.8)
Non-Hispanic	1831 (96.0)	861 (96.5)	665 (96.2)
Race, n (%)[a]			
Asian	7 (0.4)	1 (0.1)	0 (0.0)
Black or African American	86 (4.4)	29 (3.2)	23 (3.3)
Hawaiian	3 (0.2)	2 (0.2)	2 (0.3)
Native American	25 (1.3)	10 (1.1)	7 (1.0)
White	1740 (89.6)	821 (91.0)	633 (90.3)
Mixed	82 (4.2)	39 (4.3)	36 (5.1)
Education, n (%)[a]			
8th grade or less	3 (0.2)	0 (0.0)	1 (0.1)
Some high school	14 (0.8)	8 (0.9)	3 (0.4)
High school graduate	175 (10.1)	83 (9.6)	66 (9.6)
Some college	658 (38.1)	305 (35.3)	242 (35.4)
College	498 (28.9)	254 (29.4)	202 (29.5)
Postgraduate	378 (21.9)	215 (24.8)	170 (24.9)

[a]Percentage does not include missing cases.

Table 4. Round 2 participant demographics.

Variable	Baseline	3-day test retest	30-day retest
Participants (n)	704	239	51
Age (years), mean (SD)	54.5 (11.8)	54.8 (12.1)	53.7 (12.7)
Conditions, median (range)	1 (1-35)	1 (1-27)	1 (1-18)
Gender, n (%)[a]			
Male	189 (26.9)	61 (25.6)	15 (29)
Female	514 (73.1)	177 (74.4)	36 (70)
Ethnicity, n (%)[a]			
Hispanic	26 (3.9)	7 (3.0)	1 (2)
Non-Hispanic	640 (96.1)	226 (97.0)	47 (97)
Race, n (%)[a]			
Asian	3 (0.4)	1 (0.4)	1 (2)
Black or African American	53 (7.8)	13 (5.5)	3 (6)
Hawaiian	0 (0.0)	0 (0.0)	0 (0)
Native American	6 (0.9)	2 (0.8)	0 (0)
White	586 (86.3)	214 (90.7)	39 (81)
Mixed	31 (4.6)	6 (2.5)	5 (10)
Education, n (%)[a]			
8th grade or less	1 (0.2)	0 (0.0)	0 (0)
Some high school	6 (1.0)	3 (1.4)	0 (0)
High school graduate	81 (13.8)	18 (8.6)	7 (17)
Some college	225 (38.5)	94 (45.0)	15 (37)
College	160 (27.4)	55 (26.3)	12 (30)
Postgraduate	112 (19.1)	39 (18.7)	6 (15)

[a]Percentage does not include missing cases**.**

Table 5. Reliability estimates for surviving thrive scales.

Thrive scale (number of items)	Internal consistency reliability (Cronbach alpha; n=704)	Test-retest reliability (n=239)	
		Absolute ICC[a]	*P* value
Overall Health (1)	—[b]	.749	<.001
Impact of Primary Condition (1)	—	.763	<.001
Core Symptoms (5)	.815	.909	<.001
Mobility (1)	—	.898	<.001
Sleep (2)	.712	.833	<.001
Abilities (5)	.853	.912	<.001
Thriving (4)	.879	.889	<.001

[a]ICC: intraclass correlation coefficient.

[b]Not applicable.

Table 6. Ability to detect change (Pearson correlations between Thrive and comparator instruments' residualized change scores in longitudinal data, N=51).

Thrive scale item	Pearson r; *P* value										
	PHQ[a]-9 (n=704)	SF[b]-20 (n=704) General Health Item	SF-20 (n=704) Mental Health	SF-20 (n=704) Physical Functioning	SF-20 (n=704) Role Functioning	SF-20 (n=704) Health Perception	MSRS[c] (n=255)	PLM-QoL[d] (n=64) Physical	PLM-QoL (n=64) Mental	PLM-QoL (n=64) Social	ALS FRS-R[e] (n=60)
Overall Health (1 item)	—[f]	.813; <.001	—	—	—	—	—	—	—	—	—
Impact of Primary Condition (1 item)	.463; <.001	—	−.445; <.001	−.439; <.001	−.443; <.001	−.518; <.001	.452; <.001	−.573; <.001	−.492; <.001	−.477; <.001	−.477; <.001
Core Symptoms (5 items)	.750; <.001	—	−.759; <.001	−.390; <.001	−.392; <.001	−.644; <.001	.574; <.001	−.698; <.001	−.775; <.001	−.675; <.001	−.148; .26
Mobility (1 item)	—	—	—	.415; <.001	—	—	−.471; <.001	.687; <.001	—	—	.423; <.001
Sleep (2 items)	−.562; <.001	—	—	—	—	—	—	—	—	—	—
Abilities (5 items)	−.744; <.001		.708; <.001	.478; <.001	.520; <.001	.671; <.001	−.687; <.001	.791; <.001	.770; <.001	.809; <.001	.450; <.001
Thriving (4 items)	−.743; <.001	—	.780; <.001	.342; <.001	.378; <.001	.626; <.001	−.453; <.001	.639; <.001	.806; <.001	.736; <.001	.132; .32

[a]PHQ: Patient Health Questionnaire.

[b]SF: Short-Form General Health Survey.

[c]MSRS: multiple sclerosis rating scale.

[d]QoL: quality of life.

[e]ALSFRS-R: amyotrophic lateral sclerosis functional rating scale-revised.

[f]Not applicable.

Table 7. Ability to detect change (Pearson correlations between Thrive and comparator instruments' residualized change scores in longitudinal data, N=51).

Variable	PHQ[a]-9, total		PHQ-9, sleep item		SF[b]-20, general health item		SF-20, mental health		SF-20, physical functioning		SF-20, role functioning		SF-20, health perception	
	r	*P* value	r	*P* value	r	*P* value	r	*P* value	r	*P* value	r	*P* value	r	*P* value
Overall health	—[c]	—	—	—	.311	.03	—	—	—	—	—	—	—	—
Impact of primary condition	.404	.003	—	—	—	—	.352	.011	.091	.53	.099	.49	.276	.05
Core symptoms	.475	<.001	—	—	—	—	.485	<.001	.217	.13	.145	.31	.510	<.001
Mobility	—	—	—	—	—	—	—	—	.269	.06	—	—	—	—
Sleep	—	—	.496	<.001	—	—	—	—	—	—	—	—	—	—
Abilities	.190	.18	—	—	—	—	.125	.384	−.005	.97	.330	.02	.219	.12
Thriving	.356	.01	—	—	—	—	.389	.005	.027	.85	.058	.69	.041	.78

[a]PHQ: Patient Health Questionnaire.

[b]SF: Short-Form General Health Survey.

[c]Not applicable.

Table 8. Final core Thrive items.

Scale name (# of items) and item label	Item content	Response options
Overall health (1)		
Overall health	Over the last month, how has your health been?	5=Excellent; 4=Very good; 3=Good; 2=Fair; 1=Poor
Impact of primary condition (1)		
Condition impact	Over the last month, how much has your [primary condition] affected your life?	0=Not at all; 1=A little; 2=Some; 3=A lot
Core symptoms (5)		
Pain	Please rate the severity of any pain over the past month	0=None; 1=Mild; 2=Moderate; 3=Severe
Depressed mood	Please rate the severity of any depressed mood over the past month	0=None; 1=Mild; 2=Moderate; 3=Severe
Anxious mood	Please rate the severity of any anxious mood over the past month	0=None; 1=Mild; 2=Moderate; 3=Severe
Fatigue	Please rate the severity of any fatigue over the past month	0=None; 1=Mild; 2=Moderate; 3=Severe
Stress	Please rate the severity of any stress over the past month	0=None; 1=Mild; 2=Moderate; 3=Severe
Mobility (1)		
Walk	Over the last month, how well could you walk without support (such as a brace, cane, or walker)?	4=Extremely well; 3=Very well; 2=Fairly well; 1=Poorly; 0=Not at all
Sleep (2)		
Fall asleep	Over the last month, how well could you fall asleep when you wanted to?	4=Extremely well; 3=Very well; 2=Fairly well; 1=Poorly; 0=Not at all
Stay asleep	Over the last month, how well could you sleep through the night?	4=Extremely well; 3=Very well; 2=Fairly well; 1=Poorly; 0=Not at all
Abilities (5)		
Think	Over the last month, how well could you think, concentrate, and remember things?	4=Extremely well; 3=Very well; 2=Fairly well; 1=Poorly; 0=Not at all
Emotions	Over the last month, how well could you control your emotions?	4=Extremely well; 3=Very well; 2=Fairly well; 1=Poorly; 0=Not at all
Personal needs	Over the last month, how well could you take care of your personal needs?	4=Extremely well; 3=Very well; 2=Fairly well; 1=Poorly; 0=Not at all
Responsibilities	Over the last month, how well could you meet your responsibilities at work, school, or home?	4=Extremely well; 3=Very well; 2=Fairly well; 1=Poorly; 0=Not at all
Social	Over the last month, how well could you participate in your favorite social and leisure activities?	4=Extremely well; 3=Very well; 2=Fairly well; 1=Poorly; 0=Not at all
Thriving (4)		
Good	Over the last month, how often did you feel good about yourself?	3=All of the time; 2=Most of the time; 1=Some of the time; 0=None of the time
Meaning	Over the last month, how often did you find meaning in your life?	3=All of the time; 2=Most of the time; 1=Some of the time; 0=None of the time
Connect	Over the last month, how often did you feel connected to others?	3=All of the time; 2=Most of the time; 1=Some of the time; 0=None of the time
Wanted	Over the last month, how often did you feel able to live the life you wanted?	3=All of the time; 2=Most of the time; 1=Some of the time; 0=None of the time

Core Symptoms

A chi-square test demonstrated that the partial credit model (PCM [39]) fit significantly better than the more parsimonious rating scale model (RSM) [40] (P<.001). Therefore, the PCM was utilized to evaluate rating scale functioning. First, unidimensionality, item fit, and item discriminations were evaluated. A principal component analysis (PCA) on the probability scale residuals provided support for unidimensionality (see Multimedia Appendix 2). Item fit was evaluated by examining mean square infit and outfit statistics estimated by the Rasch model. Items exhibited acceptable fit

to the model [41]. Item discrimination statistics were similar, although the Pain item discriminated between persons less well than the other items (discrimination=.61; see Multimedia Appendix 2 for further details).

Andrich thresholds were ordered, providing evidence that the items' rating scales were functioning as expected [42]. Evaluation of the person-to-item map suggested adequate coverage across the latent construct (see Multimedia Appendix 2). Next, the presence of bias was evaluated via differential item function (DIF) in WINSTEPS. DIF was considered notable if the DIF contrast estimate was >1.0 in absolute value and statistically significant [43,44]. Although the presence of DIF can suggest that an item is not fair or biased, significant DIF can also indicate that the groups truly differ on the construct being measured [44]. Results did not reveal evidence of DIF for gender or race (white and nonwhite). However, results suggested the presence of DIF for the Anxious Mood item between the autoimmune relapsing and psychiatric groups, whereby this item was easier to endorse for the autoimmune relapsing group. Internal consistency was good, and stability was excellent (Table 5). Results largely provided support for convergent validity (Table 6) and ability to detect change (Table 7).

Mobility

Absolute agreement of responses to the Walk item across the 3-day test-retest period was good (Table 5). This single-item scale was related as anticipated to comparator measures, providing support for convergent validity (Table 6). A positive correlation between the Walk item's and the SF-20 Physical Functioning scale's residualized change scores (see Table 7) provide support for the Walk item's ability to detect change over time.

Sleep

The PCM did not evidence significantly better fit than the RSM, so the RSM was used to evaluate rating scale functioning. Assumptions of the model were met, and results suggested that the rating scale was performing as expected. The items did not show evidence of DIF for gender, race, or condition (autoimmune relapsing, psychiatric, or neurodegenerative). Internal consistency was acceptable, and stability was good (Table 5). Due to shared content, a Pearson correlation between the PHQ-9 and Sleep scale was calculated to evaluate convergent validity of the Sleep scale. Results provided support for convergent validity. Finally, the positive correlation between Sleep and the PHQ-9 Sleep item's residualized change scores provides evidence of ability to detect change over time (Table 7).

Abilities

Of the Abilities items, 1 ("Over the last month, how well could you live the life you wanted to live?") was removed because of conceptual redundancy with another item ("Over the last month, how often did you feel able to live life you wanted?").

A global chi-square fit test demonstrated that the PCM fit significantly better than the RSM (P<.001). Therefore, the PCM was utilized to evaluate rating scale functioning. Results from a PCA on the probability scale residuals provided support for unidimensionality. The items exhibited acceptable item fit and similar item discriminations.

The items' rating scales were functioning as expected, and examination of the person-to-item map suggests adequate coverage. The items did not evidence DIF for gender or race. However, results suggested the presence of DIF for the Think item between the neurodegenerative group and the autoimmune group, whereby this item was easier to endorse for patients with neurodegenerative conditions. Internal consistency was good, and stability was excellent (Table 5). Pearson correlations provided support for convergent validity (Table 6). Results largely provided support for ability to detect change (Table 7). However, the residualized change scores for Abilities and SF-20 Physical Functioning evidenced a near-zero correlation. Evaluation of the SF-20 Physical Functioning composite reveals that items reflect physical mobility and ability to engage in vigorous physical activity (eg, lifting heavy objects, running, walking, walking uphill, bending, etc). Therefore, it is not surprising that change in the 2 scales over time were not related.

Thriving

A chi-square test demonstrated that the PCM fit significantly better than the RSM (P<.001). Therefore, the PCM was utilized to evaluate rating scale functioning. Of the items, 1 ("Over the last month, how often did you stick to the health habits you wanted to?") was removed because of poor model fit and discrimination (.25). Following removal of this item, another item ("Over the last month, how often did you feel able to take charge of your health?") was also removed because of poor discrimination (.67). The remaining items evidenced acceptable levels of fit [41] and discrimination [45], as well as unidimensionality based on results from a PCA of the probability scale residuals. Results suggested that the items' rating scales were functioning as expected.

Next, for purposes of reducing the scale length, the research team utilized theoretical (review of item content) and empirical (person-to-item map, interitem correlations) rationale to identify items for removal. As a result, 4 additional Thriving items were removed ("Over the last month, how often did you feel confident that you could handle your life?," "Over the last month, how often did you see yourself as a worthwhile person?," "Over the last month, how often did you feel effective?," and "Over the last month, how often did you feel you were thriving?"). Removing these items did not result in substantial loss of reliability (from a person reliability coefficient of .92 to a person reliability coefficient of .86). The remaining 4 items evidenced good person-to-item coverage and did not evidence DIF for gender, race, or condition.

Internal consistency and stability were good (Table 5). Pearson correlations largely provided support for convergent validity (Table 6) and ability to detect change (Table 7). The PHQ-9 and SF-20 Mental Health scales' residualized change scores were significantly related to change in Thriving scores over the 30-day period, whereas near-zero correlations were observed between change in Thriving and the remaining SF-20 scales.

Scoring

Scores for the multi-item scales (Core Symptoms, Sleep, Abilities, and Thriving) are calculated by taking the average of the items. Whether or not scores are calculated when data are missing depends on how the instrument is being used. For example, PatientsLikeMe members can complete Thrive on a monthly basis to track their functioning, and composites for the Thrive domains can be calculated with missing data so long as 80% of items are completed for each domain. Of course, calculating a score with missing items can increase measurement error. Therefore, whenever possible, patients should be encouraged to answer as many items as they feel comfortable answering.

Discussion

Principal Findings

PROs have the potential to move the locus of control in health care from institutions and professionals to patients themselves by enabling digital health tools that track and predict outcomes, alert their health team, support shared decision making, enable learning from their peer group, underpin systematic self-experimentation, and let them continually participate in research [46]. Building tools that motivate users to *want* to come back and enter data requires PROs that pay as much respect to principles of user design and user experience as they do to psychometric validity [47]. This is a new challenge for a field more used to designing instruments on paper for researchers to administer in blinded clinical trials, but it is one we will have to address to help fight the law of attrition [48] and gather sufficient data to understand their disease and make better decisions as part of a learning health system that is *by the people, for the people* [49].

Following established best practice for instrument development [16], we have demonstrated that a novel set of PRO items (Thrive Core Items) can adequately describe the key domains of HRQoL in adults with chronic illness in a way that is positive and aspirational. Detailed psychometric analysis was used to refine the instrument to reduce burden and redundancy, and comparison with validated generic and condition-specific legacy PRO measures suggest an acceptable degree of agreement. Many PROs used in research and clinical practice today focus almost exclusively on *how bad a life* patients are living as a result of disease. Given that nearly all chronic health conditions are incurable and progressive, they serve only to document an individual's descent into infirmity. Tools that encourage a positive mindset and support goal-setting to thrive despite symptoms and disability may well be important in encouraging patients to live their best life by seeking pleasure, engagement, and meaning [50].

During our interviews, patients consistently described *disease* only as a problem to be managed, *health* as the overall state of their bodies and minds, and *thriving* as living the life they wanted to live. Of the participants, 1 remarked:

> *Health incorporates disease but is bigger. Health is the ability to enjoy life with minimal impact from your conditions. It's feeling good about life and who you are. Thriving is even more than health...it's looking forward to each day with desire...and feeling that life is good.*

After reviewing the items, most participants interviewed agreed that the Thrive Core items regarding meaning, connectedness to others, self-esteem, and coping were best at reflecting what thriving meant to them.

Advantages of Thrive for Digital Health

Thrive contains a number of features designed to make it appealing for use in digital health. Using consistent items across multiple conditions is supportive of patients with multiple comorbidities. For example, a patient living with both PD and MDD only needs to complete information about shared domains (such as ability to sleep) once. By contrast, in our previous PRO model, a patient would have been asked to complete not only a Parkinson-specific measure (the PDRS) but also a mood-specific measure (the mood map) and a generic HR-QoL measure (PLM-QoL), with a number of additional symptoms. The burden of this battery of instruments (100 items with 3 different recall periods, 5 different response scales, and some 3252 words to read) is dramatically reduced by Thrive (19 core items plus 22 condition-specific questions [41 total] in 924 words across consistent response scales and recall periods). Question stems and response options are short and consistent, being optimal for use on mobile displays. When deployed on PatientsLikeMe, users have the option to respond "stop asking me this" for each item, which may be particularly useful for members with quadriplegia whose condition will not improve, those who feel emotionally *triggered* by certain questions, or who are in good physical health but only want to track mood or other psychological symptoms. Although fewer than 1% of real-world users choose to switch off an item (internal data), interviewees felt this option offered a greater sense of control over their own experience rather than attempting to skip an item or enter false data to skip to the next screen.

Limitations

This study was subject to a number of limitations. Although the overall number of participants recruited was relatively large, it was a convenience sample from users of an online health community, had only a 9.5% completion rate from those invited, and there was a bias for participants to be more likely to be female and well-educated. There was significant attrition in both rounds of the 3-day retest and 30-day follow-up, which limited our ability to detect minimally important differences and may limit generalizability. Our sample was limited to English-speaking participants residing in the United States with a handful of chronic health conditions. All this limits generalizability to other populations and should be tested further. A larger, prospective, longitudinal study over a longer time course would have been preferable to establish minimally important differences and sensitivity to change. Although Thrive will be deployed with multiple items relating to both the *Impact of Primary Condition* and additional *Impact of additional conditions* related to their comorbidities, this study only asked about a single condition. This may have obscured the impact of important comorbidities.

The number of cognitive interviews conducted was a *total* of 12 participants; arguably we might have interviewed 12 patients for *each* of the 9 condition groups represented in the sample [51]. However, as we were developing a measure for chronic illness more generally rather than specific conditions, this was considered adequate, and both of our interviewers felt we had achieved *saturation* [52]. Interviewing over 100 participants was also considered infeasible in the time and budget allowed.

All participant data were self-reported rather than being independently validated, though previous studies suggest a high degree of agreement between patient self-report of diagnosis and confirmation via, for example, insurance claims [53]. Some of the condition-specific comparator measures used on PatientsLikeMe, and by extension, this study, were unvalidated—they were tested to further our plans to remove them from our online community but do not provide as robust tests as a validated measure would have achieved. However, the use of the widely used SF-20 and PHQ-9 make up for this shortcoming to some degree. Owing to the number of conditions and comparator measures, our reporting of convergent validity was necessarily more simplistic than desirable. Small samples for condition-specific measures such as the PLM-QoL, MSRS, and ALSFRS-R relative to the PHQ-9 or SF-20 may have obscured the strength of relationships for comparative validity in the former. Next steps for validation include deployment of the Core Thrive Items in a more representative sample of US citizens and testing of disease-specific versions of Thrive in clinical settings alongside ClinROs and objective measures such as blood tests.

Modularity for Expansion and Future Research

Analysis of comparative validity suggests that although there are moderate-strong correlations with overlapping domains from other instruments, it is unlikely that the Core items represent complete *coverage* of all the important domains for every condition. For example, there were only moderate correlations between the Mobility and Abilities scales with the ALSFRS-R [12], and clinical experience tells us that a measure that fails to take speech, swallowing, feeding, or breathing ability into account would be missing key data for understanding patients and their disease.

Work is already in progress to describe the development of condition-specific item banks that can be interspersed with the Thrive Core Items (Figure 2). Review of existing PROs, the clinical literature, and the patient-added symptoms of existing PatientsLikeMe users have been used to add additional domains such as *tremor* as a symptom in Parkinson disease, or *breathing* as an ability in ALS, for instance. Future studies will describe condition-specific validation of Thrive-Condition Instruments such as *Thrive-ALS* against legacy measures such as the ALSFRS-R [12] in more detail, with the addition of clinical and other objective biomarkers where possible. Such clinical work will also be useful in establishing minimally important differences for changes on different Thrive subscores over time and in response to treatment. The Thrive Core Items are available to members of PatientsLikeMe.com as a *MonthlyMe* interview, and although the psychometric validation described herein is probably sufficient to support patient self-tracking and visualization of individual items and subscores (eg, showing how an individual compares to a group of patients like them, or showing relationships between different variables), further condition-specific work is needed to confirm the tool's validity for clinical management or proving a treatment effect in clinical trial.

Figure 2. Sample additional items for 2 conditions based on health care professional review. MS: multiple sclerosis; PTSD: posttraumatic stress disorder.

	Core instrument items	Sample additional items for MS	Sample additional items for PTSD
Symptom severity items	▪ Anxious mood ▪ Depressed mood ▪ Fatigue ▪ Pain ▪ Stress	+ Balance problems + Bladder problems + Bowel problems + Numbness and tingling + Stiffness/spasticity + Tremor(s) + Visual disturbance	+ Aggressive behavior + Hypervigilance + Increased appetite + Increased need for sleep + Inability to experience pleasure + Intrusive thoughts or impulses + Irritability
Ability items	▪ Control emotions ▪ Fall asleep ▪ Manage personal needs ▪ Meet responsibilities ▪ Participate socially ▪ Stay asleep ▪ Think ▪ Walk	+ See + Speak + Swallow + Use your hands	+ Manage your response to triggers
Thriving items	▪ Feel connected with others ▪ Feel good about self ▪ Find meaning in life ▪ Live the life you want		+ Feel motivated to do everyday things + Feel hopeful

Work with partners may also involve translation into other languages (such as Mandarin Chinese) and deployment through mobile messaging platforms (such as WeChat) as part of wellness apps. Finally, future work will consider the role of treatment side effects and treatment burden as key aspects of thriving despite illness [54]. We offer the use of Thrive under Creative Commons Attribution ShareAlike International (CCSA-4.0) so that others can deploy and adapt it in their work to measure what matters most to people. Although there is always the risk in taking this approach that some users may use the instrument inappropriately (eg, by adding poor quality items or mistranslating into other languages), we believe the tangible benefits of making an instrument freely available outweigh the theoretical harms.

Conclusions

Validation is a continuous and iterative process. This study describing the development and testing of the Thrive Core Set items is the first step on a path that includes replacing all the PROs on PatientsLikeMe, testing against putative biomarkers of disease progression, and deployment on third party digital health platforms. We hope Thrive will be a key resource in the digitization of human health to improve longevity and well-being for all.

Acknowledgments

The authors are grateful to the many participants who took part in their validation efforts, and to colleagues at PatientsLikeMe for their support of this work. The PatientsLikeMe Research Team has received research funding (including conference support and consulting fees) from Abbvie, Accorda, Actelion, Alexion, Amgen, AstraZeneca, Avanir, Biogen, Boehringer Ingelheim, Celgene, EMD, Genentech, Genzyme, Janssen, Johnson & Johnson, Merck, Neuraltus, Novartis, Otsuka, Permobil, Pfizer, Sanofi, Shire, Takeda, Teva, and UCB. The PatientsLikeMe R&D team has received research grant funding from Kaiser Permanente, the Robert Wood Johnson Foundation, Sage Bionetworks, The AKU Society, and the University of Maryland. PW has received speaker fees from Bayer and honoraria from Roche, ARISLA, AMIA, IMI, PSI, and the BMJ.

Conflicts of Interest

PW, MH, and JH are employees of PatientsLikeMe and hold stock options in the company. KG, SM, and RB conducted this work as paid consultants to PatientsLikeMe. PW is an associate editor at the *Journal of Medical Internet Research* and is on the editorial boards of *BMJ*, *BMC Medicine*, and *Digital Biomarkers*.

Multimedia Appendix 1

Consolidated Standards of Reporting Trials flow diagrams.

[PDF File (Adobe PDF File), 68KB - jmir_v21i1e12075_app1.pdf]

Multimedia Appendix 2

Detailed round 1 psychometric analysis.

[PDF File (Adobe PDF File), 454KB - jmir_v21i1e12075_app2.pdf]

References

1. Kluetz PG, O'Connor DJ, Soltys K. Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. Lancet Oncol 2018 May;19(5):e267-e274. [doi: 10.1016/S1470-2045(18)30097-4] [Medline: 29726391]
2. Lowe MM, Blaser DA, Cone L, Arcona S, Ko J, Sasane R, et al. Increasing patient involvement in drug development. Value Health 2016;19(6):869-878 [FREE Full text] [doi: 10.1016/j.jval.2016.04.009] [Medline: 27712716]
3. Acquadro C, Berzon R, Dubois D, Leidy N, Marquis P, Revicki D, PRO Harmonization Group. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. Value Health 2003;6(5):522-531 [FREE Full text] [doi: 10.1046/j.1524-4733.2003.65309.x] [Medline: 14627058]
4. Emery MP, Perrier LL, Acquadro C. Patient-reported outcome and quality of life instruments database (PROQOLID): frequently asked questions. Health Qual Life Outcomes 2005 Mar 8;3:12 [FREE Full text] [doi: 10.1186/1477-7525-3-12] [Medline: 15755325]
5. Perret C, Perrier L, Castex M, Acquadro C. Proqolid database: evolution of content, structure, and functionalities (2012-2016)-integration in eprovide, a new online platform dedicated to clinical outcome assessment (Coa) research. Value Health 2016 Nov;19(7):A369. [doi: 10.1016/j.jval.2016.09.131]
6. Dobrozsi S, Panepinto J. Patient-reported outcomes in clinical practice. Hematology Am Soc Hematol Educ Program 2015;2015:501-506. [doi: 10.1182/asheducation-2015.1.501] [Medline: 26637765]

7. Howell D, Molloy S, Wilkinson K, Green E, Orchard K, Wang K, et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. Ann Oncol 2015 Sep;26(9):1846-1858 [FREE Full text] [doi: 10.1093/annonc/mdv181] [Medline: 25888610]
8. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
9. Omer T. Empowered citizen 'health hackers' who are not waiting. BMC Med 2016 Dec 17;14(1):118 [FREE Full text] [doi: 10.1186/s12916-016-0670-y] [Medline: 27530970]
10. Wicks P, Vaughan T, Heywood J. Subjects no more: what happens when trial participants realize they hold the power? Br Med J 2014 Jan 28;348:g368 [FREE Full text] [doi: 10.1136/bmj.g368] [Medline: 24472779]
11. Wicks P, Massagli MP, Wolf C, Heywood J. Measuring function in advanced ALS: validation of ALSFRS-EX extension items. Eur J Neurol 2009 Mar;16(3):353-359. [doi: 10.1111/j.1468-1331.2008.02434.x] [Medline: 19364363]
12. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS study group (phase III). J Neurol Sci 1999 Oct 31;169(1-2):13-21. [Medline: 10540002]
13. Wicks P. Commentary: measuring what matters: the case for patient generated PROMS. Br Med J 2015 Feb 10;350:h54. [doi: 10.1136/bmj.h54] [Medline: 25670198]
14. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
15. Maier A, Holm T, Wicks P, Steinfurth L, Linke P, Münch C, et al. Online assessment of ALS functional rating scale compares well to in-clinic evaluation: a prospective trial. Amyotroph Lateral Scler 2012 Feb;13(2):210-216 [FREE Full text] [doi: 10.3109/17482968.2011.633268] [Medline: 22292842]
16. US Department of Health and Human Services FDA Center for Drug Evaluation and Research. US Department of Health and Human Services FDA Center for Drug Evaluation and Research. 2006 Oct 11. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims URL: https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf [accessed 2019-01-17] [WebCite Cache ID 75UPuEnhO]
17. Cano SJ, Hobart JC. The problem with health measurement. Patient Prefer Adherence 2011;5:279-290 [FREE Full text] [doi: 10.2147/PPA.S14399] [Medline: 21792300]
18. Wicks P, Mack Thorley E, Simacek K, Curran C, Emmas C. Scaling PatientsLikeMe via a "Generalized Platform" for members with chronic illness: web-based survey study of benefits arising. J Med Internet Res 2018 May 7;20(5):e175 [FREE Full text] [doi: 10.2196/jmir.9909] [Medline: 29735472]
19. Ferrans CE, Zerwic JJ, Wilbur JE, Larson JL. Conceptual model of health-related quality of life. J Nurs Scholarsh 2005;37(4):336-342. [Medline: 16396406]
20. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. J Am Med Assoc 1995 Jan 4;273(1):59-65. [doi: 10.1001/jama.1995.03520250075037] [Medline: 7996652]
21. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]
22. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
23. International Classification Of Functioning, Disability And Health. Geneva: World Health Organization; 2001. URL: http://www.who.int/sorry/ [accessed 2019-01-17] [WebCite Cache ID 75UQMIa3j]
24. Tiplady B, Byrom B, editors. ePRO: Electronic Solutions For Patient-reported Data. Farnham, UK: Routledge; 2010.
25. 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]
26. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002 Sep 1;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
27. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the brief Patient Health Questionnaire mood scale (PHQ-9) in the general population. Gen Hosp Psychiatry 2006;28(1):71-77. [doi: 10.1016/j.genhosppsych.2005.07.003] [Medline: 16377369]
28. Löwe B, Kroenke K, Herzog W, Gräfe K. Measuring depression outcome with a brief self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). J Affect Disord 2004 Jul;81(1):61-66. [doi: 10.1016/S0165-0327(03)00198-8] [Medline: 15183601]
29. Ware AE, Sherbourne CD, Davies AR. Developing and testing the MOS 20-Item short-form health survey. In: Stewart AL, Ware Jr AE, editors. Measuring Functioning And Well-being: The Medical Outcomes Study Approach. Durham, NC: Duke University Press Books; 1992:277-290.
30. Stewart AL, Hays RD, Ware Jr JE. The MOS short-form general health survey. Reliability and validity in a patient population. Med Care 1988 Jul;26(7):724-735. [Medline: 3393032]

31. Sharrack B, Hughes RA. Scale development and Guy's Neurological Disability Scale. J Neurol 1999 Mar;246(3):226. [Medline: 10323324]
32. Bove R, Secor E, Healy BC, Musallam A, Vaughan T, Glanz BI, et al. Evaluation of an online platform for multiple sclerosis research: patient description, validation of severity scale, and exploration of BMI effects on disease course. PLoS One 2013;8(3):e59707 [FREE Full text] [doi: 10.1371/journal.pone.0059707] [Medline: 23527256]
33. Kaufmann P, Levy G, Thompson JL, Delbene ML, Battista V, Gordon PH, et al. The ALSFRSr predicts survival time in an ALS clinic population. Neurology 2005 Jan 11;64(1):38-43. [doi: 10.1212/01.WNL.0000148648.38313.64] [Medline: 15642901]
34. Kollewe K, Mauss U, Krampfl K, Petri S, Dengler R, Mohammadi B. ALSFRS-R score and its ratio: a useful predictor for ALS-progression. J Neurol Sci 2008 Dec 15;275(1-2):69-73. [doi: 10.1016/j.jns.2008.07.016] [Medline: 18721928]
35. Slawsky KA, Massagli MP, Wicks P. PRM30 a comparison of the Patientslikeme Quality of Life Questionnaire (PLMQOL) with the Rand SF-36. Value Health 2011;14(7):A426. [doi: 10.1016/j.jval.2011.08.1061]
36. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
37. Bond TG, Fox CM. Applying The Rasch Model: Fundamental Measurement in the Human Sciences. New York: Routledge; 2015.
38. Furr RM. Psychometrics: An Introduction. Thousand Oaks: Sage Publications; 2018.
39. Masters GN. A Rasch model for partial credit scoring. Psychometrika 1982 Jun;47(2):149-174. [doi: 10.1007/BF02296272]
40. Andrich D. Application of a psychometric rating model to ordered categories which are scored with successive integers. Appl Psychol Meas 1978 Oct 1;2(4):581-594. [doi: 10.1177/014662167800200413]
41. Linacre JM. Winsteps. A User's Guide to WINSTEPS MINISTEP Rasch-Model Computer Programs URL: https://www.winsteps.com/winman/copyright.htm [accessed 2019-01-16] [WebCite Cache ID 75SxgAJi9]
42. Embretson SE, Reise SP. Item Response Theory for Psychologists. London: Psychology Press; 2000.
43. Clauser BE, Mazor KM. Using statistical procedures to identify differentially functioning test items. Educ Meas 2005 Oct 25;17(1):31-44. [doi: 10.1111/j.1745-3992.1998.tb00619.x]
44. Zieky M. Educational Testing Service. A DIF Primer URL: https://www.ets.org/s/praxis/pdf/dif_primer.pdf [accessed 2019-01-23] [WebCite Cache ID 75deU8Zr9]
45. Kelley T, Ebel R, Linacre JM. Institute for Objective Measurement. 2002. Item discrimination indices URL: https://www.rasch.org/rmt/rmt163a.htm [accessed 2019-01-16] [WebCite Cache ID 75Sxqll0f]
46. Wicks P, Hotopf M, Narayan VA, Basch E, Weatherall J, Gray M. It's a long shot, but it just might work! Perspectives on the future of medicine. BMC Med 2016 Nov 7;14(1):176 [FREE Full text] [doi: 10.1186/s12916-016-0727-y] [Medline: 27817747]
47. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015 Jun;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]
48. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
49. Okun S, Goodwin K. Building a learning health community: by the people, for the people. Learn Health Syst 2017 Jun 6;1(3):e10028. [doi: 10.1002/lrh2.10028]
50. Duckworth AL, Steen TA, Seligman ME. Positive psychology in clinical practice. Annu Rev Clin Psychol 2005;1:629-651. [doi: 10.1146/annurev.clinpsy.1.102803.144154] [Medline: 17716102]
51. Lasch KE, Marquis P, Vigneux M, Abetz L, Arnould B, Bayliss M, et al. PRO development: rigorous qualitative research as the crucial foundation. Qual Life Res 2010 Oct;19(8):1087-1096 [FREE Full text] [doi: 10.1007/s11136-010-9677-6] [Medline: 20512662]
52. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. Qual Life Res 2009 Nov;18(9):1263-1278. [doi: 10.1007/s11136-009-9540-9] [Medline: 19784865]
53. Eichler GS, Cochin E, Han J, Hu S, Vaughan TE, Wicks P, et al. Exploring concordance of patient-reported information on PatientsLikeMe and medical claims data at the patient level. J Med Internet Res 2016;18(5):e110 [FREE Full text] [doi: 10.2196/jmir.5130] [Medline: 27174602]
54. Tran VT, Harrington M, Montori VM, Barnes C, Wicks P, Ravaud P. Adaptation and validation of the Treatment Burden Questionnaire (TBQ) in English using an internet platform. BMC Med 2014 Jul 2;12:109 [FREE Full text] [doi: 10.1186/1741-7015-12-109] [Medline: 24989988]

Abbreviations

ALS: amyotrophic lateral sclerosis
ALSFRS-R: ALS functional rating scale revised
ClinRO: clinician-reported outcome
DIF: differential item function
GAD: generalized anxiety disorder

HRQoL: health-related quality of life
MDD: major depressive disorder
MS: multiple sclerosis
MSRS: multiple sclerosis rating scale
PCA: principal component analysis
PCM: partial credit model
PD: Parkinson disease
PDRS: Parkinson's disease rating scale
PHQ: Patient Health Questionnaire
PRO: patient-reported outcome
PTSD: posttraumatic stress disorder
RSM: rating scale model
SF: Short-Form General Health Survey
SLE: systemic lupus erythematosus

Edited by G Eysenbach; submitted 30.08.18; peer-reviewed by A Mohammadabadi, SS Alavi, LT Wu, J Boehnke; comments to author 21.09.18; revised version received 02.11.18; accepted 10.11.18; published 25.01.19.

Please cite as:
Wicks P, McCaffrey S, Goodwin K, Black R, Hoole M, Heywood J
A Modular Health-Related Quality of Life Instrument for Electronic Assessment and Treatment Monitoring: Web-Based Development and Psychometric Validation of Core Thrive Items
J Med Internet Res 2019;21(1):e12075
URL: http://www.jmir.org/2019/1/e12075/
doi:10.2196/12075
PMID:30681962

Original Paper

Improving the Efficacy of Cognitive Training for Digital Mental Health Interventions Through Avatar Customization: Crowdsourced Quasi-Experimental Study

Max Valentin Birk[1*], PhD; Regan Lee Mandryk[1*], PhD

University of Saskatchewan, Department of Computer Science, Saskatoon, SK, Canada

[*]all authors contributed equally

Corresponding Author:
Max Valentin Birk, PhD
University of Saskatchewan
Department of Computer Science
110 Science Place
Saskatoon, SK, S7N 5C9
Canada
Phone: 1 3069662327
Email: max.birk@usask.ca

Abstract

Background: The success of internet-based mental health interventions in practice—that is, in the wild—depends on the uptake and retention of the application and the user's focused attention in the moment of use. Incorporating game-based motivational design into digital interventions delivered in the wild has been shown to increase uptake and retention in internet-based training; however, there are outstanding questions about the potential of game-based motivational strategies to increase engagement with a task in the moment of use and the effect on intervention efficacy.

Objective: Designers of internet-based interventions need to know whether game-based motivational design strategies can increase in-the-moment engagement and thus improve digital interventions. The aim of this study was to investigate the effects of 1 motivational design strategy (avatar customization) in an example mental health intervention (computerized cognitive training for attention bias modification).

Methods: We assigned 317 participants to either a customized avatar or an assigned avatar condition. After measuring state anxiety (State-Trait Anxiety Inventory), we randomly assigned half of the participants in each condition to either an attentional retraining condition (Attention Bias Modification Training) or a control condition. After training, participants were exposed to a negative mood induction using images with strong negative valance (International Affective Picture System), after which we measured state anxiety again.

Results: Avatar customization decreased posttraining state anxiety when controlling for baseline state anxiety for those in the attentional retraining condition; however, those who did not train experienced decreased resilience to the negative mood induction ($F_{1,252}$=6.86, P=.009, η_p^2=.027). This interaction effect suggests that customization increased task engagement with the intervention in the moment of use. Avatar customization also increased avatar identification ($F_{5,252}$=12.46, P<.001, R^2=.23), regardless of condition ($F_{1,252}$=.79, P=.38). Avatar identification reduced anxiety after the negative mood induction for participants who underwent training but increased poststimulus anxiety for participants who did not undergo training, further suggesting that customization increases engagement in the task ($F_{1,252}$=6.19, P=.01). The beneficial effect of avatar customization on training was driven by participants who were low in their basic satisfaction of relatedness ($F_{10,248}$=18.5, P<.001, R^2=.43), which is important because these are the participants who are most likely in need of digital interventions for mental health.

Conclusions: Our results suggest that applying motivational design—specifically avatar customization—is a viable strategy to increase engagement and subsequently training efficacy in a computerized cognitive task.

(J Med Internet Res 2019;21(1):e10133) doi:10.2196/10133

KEYWORDS

cognitive therapy; computer-assisted therapy; video games; attentional bias; cognitive bias; motivation

Introduction

Background

Internet-based mental health interventions are necessary to address a growing gap between an increased demand for treatment of mental health issues and the capacity of traditional therapeutic approaches to meet this growing demand [1]. It has been argued that internet-based mental health interventions have some benefits over traditional approaches, especially related to the increased accessibility of treatment (eg, because of geographical constraints of living in remote areas with access to clinical treatment), the ability to scale (eg, to seamlessly address the growing demand of people who could benefit from treatment), the ease of access (eg, through deployment of smartphones or websites), and the broad appeal (eg, to treat subclinical populations who may not qualify for treatment in health care systems that are already stretched in meeting the needs of clinical populations) [1-3].

To be successful, an internet-based mental health intervention requires *good intervention design* as evidenced by clinical efficacy. There are several examples of internet-based mental health applications that have demonstrated treatment effects [4,5] in randomized controlled trials (RCTs) [6-8]. However, because these mental health interventions are delivered digitally and often outside a laboratory or clinical context, success is not only defined by efficacy in *research* but also defined by efficacy in *practice*, which depends on validated implementation models and a focus on external validity—that is, demonstrated success in the context of intended use [4,5,9]. Success in practice requires *uptake* (ie, people have to use the applications), *retention* (ie, people have to use the applications over a long enough term to experience benefits), and demonstrated treatment effects under conditions of *practice* (ie, the clinical efficacy must translate into less controlled environments with multiple competing demands for a patient's attention and time). These elements of demonstrated success in practice have been sources of failure for internet-based mental health interventions that have been effective in RCTs but are compromised by low uptake (ie, recruitment challenges [10]), attention (ie, failure to effectively engage participants in the moment of intervention use [11]), and retention (ie, failure in adherence over the medium term [12]) when delivered *in the wild*. Researchers have recently promoted a need for demonstrated success of treatment efficacy in both research and practice [5,9], and the failure to close the research-practice gap has prompted suggestions that researchers still have a lot to learn about how to implement these types of interventions in the wild [5].

Uptake, Attention, and Retention

Researchers have argued that improving the in the wild success of internet-based mental health interventions can be facilitated by improving the *user experience design* [5]. Internet-based applications built for the purposes of leisure or enjoyment also depend on user engagement—that is, good uptake, attention, and retention—because the microtransaction model of monetization for the companies that build these applications requires that end users chose the application (uptake), engage with it (attention), and continue to use it (retention). Various interaction design strategies have been employed by leisure application designers to improve user engagement, including gamification—that is, the use of game-based elements in nongame contexts [13], the inclusion of extrinsic rewards [14], or leveraging social pressure to engage [15]. On the basis of increasing a user's motivation by increasing their enjoyment of and invested effort in engaging with an application, these types of interaction design strategies are part of a growing field of research on motivational design for interactive technologies that have also been recently applied to the context of digital intervention design.

For example, applying motivational design principles has been shown to foster engagement with a 12-week neurofeedback treatment for children with fetal alcohol spectrum disorder [16], to increase motivation to use a stroke rehabilitation program [17], and to improve enjoyment of a social physical activity intervention for children with cerebral palsy [18]. Furthermore, a series of studies employing avatar customization (a motivational design strategy built on increased autonomy and identification) was shown to increase enjoyment of and effort invested in a game for training [19], to increase the time spent (free choice) in a game [20], and to combat attrition in an internet-based daily breathing exercise deployed over 3 weeks [21].

Engagement in the Moment of Use

Previous work has suggested that motivational design principles could be employed to help close the research-practice gap for the design of internet-based mental health interventions by improving user engagement with the application [2]; however, prior work that led to these ideas has measured success through 2 approaches: first, by focusing on subjective measures of motivation, such as increased enjoyment of the intervention or perception of effort invested in the task (eg, [19]), and second, through metrics that operationalize intervention usage statistics, such as the time spent in treatment or number of returning sessions (eg, [21]). However, there is a third approach to characterizing success of motivational design as applied to digital interventions that has been underserved, that is, *metrics that characterize increased engagement with the task in the moment of use*. As researchers, we must differentiate between motivational design that results in greater exposure to treatment (ie, more time spent in training, more adherence)—which should result in improved efficacy through mere exposure and motivational design that fosters task engagement in the moment (eg, greater attention and focus and reduced response to distraction)—which should improve treatment efficacy without an accompanying increase in exposure.

In this paper, we employ 1 motivational design strategy (avatar customization) in an example mental health intervention (computerized cognitive training, CCT) to demonstrate that motivational design principles can not only improve exposure

to treatment through greater uptake and retention but that they can improve focused engagement in the moment of intervention use.

Computerized Cognitive Training as a Mental Health Intervention

CCT is one approach to intervention design with a focus on improving specific aspects of cognition. Feasibility studies on CCT have been shown to improve memory, self-control, reasoning, attention bias, and processing speed [22]; CCT has successfully been applied in clinical research to combat mental illness and cognitive deficits in disorders such as dementia [23], depression [24-26], neurodegenerative diseases [26], attention-deficit/hyperactivity disorder (ADHD) [27], and brain injury [28]. The most common CCT tasks are the Go/No-Go task [29], memory training [30], and the Attention Bias Modification Training (ABMT) task [31]. For example, Go/No-Go tasks require participants to inhibit responses under changing conditions (eg, in a fast-paced task, press L when a red box appears, but inhibit pressing L when a green box appears). Go/No-Go paradigms have been applied to the training of executive function [32], and research suggests that the paradigm can improve hyperactivity symptoms for children with ADHD [33] and reduce undesirable food intake when applied in the context of eating behavior [34].

Another approach to cognitive retraining is through attention bias modification [35], which exposes participants simultaneously to negative and neutral stimuli but reinforces an attentional shift toward neutral stimuli by presenting target probes behind only the neutral stimuli. ABMT has been shown to be an effective technique to shift a participant's attention away from negative stimuli, to decrease self-reported anxiety, and to decrease the response to negative stimuli [31,36,37]. Although lab-based ABMT training has been shown to be effective [36], internet-based ABMT has generally not been shown to be effective, suggesting that the training task itself might require adjustment before dissemination over the internet [38]. In the case of training tasks that require the full attention of the patient, the lack of control over distractions and attention in the environment present in internet-based interventions may compromise treatment efficacy when delivered in the wild: for CCT to be fully effective, participants need to be vigilant, psychologically present, and engaged in the task.

CCTs have shown moderate-to-large effect sizes for improving attention, working memory, and global functioning [24]. However, to show effects in the wild, CCTs need to be designed in ways that maximize user engagement in the moment. Being inattentive, unfocused, or distracted will diminish the efficacy of attention-based training [39,40]. When CCT is applied in studies or during a session with a therapist, participants are externally regulated to focus on the task; however, this external regulation is drastically lessened when people engage in cognitive training during a commute, at home, or while they have a few minutes waiting in line. To support the success of internet-based mental health interventions in the wild, researchers need to ask, *how can we increase in-the-moment engagement to compensate for inattentiveness and distraction in the wild and subsequently improve training efficacy?*

Engagement

Although internet-based mental health applications have increased the accessibility of treatment, their use still requires participant engagement and effort [11]. Theories of motivation provide guidance on how to design applications to maximize engagement. Self-Determination Theory (SDT), a well-established theory of human motivation [41], postulates that competence, relatedness, and autonomy are 3 predictors of motivation, which is expressed in terms of enjoyment, engagement, and effort. Competence—experiencing mastery over a task, autonomy—volitionally engaging in a task, and relatedness— experiencing connectedness to others—predict engagement and have been shown to be positively related to treatment outcome [11]. For example, clients who engage in therapy out of personal choice, that is, autonomy, are more likely to benefit from therapy. Increased volitional engagement has been shown to increase adherence and treatment efficacy [42].

Designing for Engagement

Multiple strategies have been applied to foster volitional engagement in a digital context. For example, gamification—the application of game elements in a nongame context [13]—has successfully been applied to increase volitional engagement in a variety of contexts [43]. Game-based training has been shown to improve working memory capacity [44], task switching ability (Ibid), visual short-term memory (Ibid), verbal reasoning (Ibid; [45]), visuospatial reasoning [46,47], response selection [48], visual attention [49,50], reaction time [51,52], and choice reaction time [53,54].

In a similar vein, persuasive technology uses strategies to bring about change by shaping or reinforcing behaviors or attitudes [55]. Two strategies commonly employed in persuasive technologies are personalization [56,57]—which is system-initiated tailoring that offers content or services personalized to an individual—and customization [56,58]—which is a system that supports user-initiated tailoring of content or services. Sundar et al [59] argue that although personalization will increase the relevance of content for individuals using an interactive system, customization yields systems and content that are not only relevant but also boost the agency and self-determination of the individual because it is they themselves who perform the tailoring. Personalization and customization have both been discussed as techniques to increase long-term engagement with internet-based interventions for mental health [21,60].

Customization fosters autonomy, a sense of control, and a sense of identity, making the person feel relevant in the context of their interaction [59]. As previously described, a series of studies employing avatar customization (which facilitated avatar identification) showed increased enjoyment of and effort invested in a game for training [19], increased the time spent (free choice) in a game [20], and combatted attrition in an internet-based daily breathing exercise deployed over 3 weeks [21]. Together, these studies show improvements in subjectively measured motivation and objectively measured motivation in terms of the time spent engaged with the application and number of return sessions in a 3-week intervention; however, there was

no attempt to demonstrate improved treatment efficacy as a result of the increased engagement with the application in the moment of use. There was no differentiation between how the motivational design resulted in greater exposure to treatment and how the motivational design fostered task engagement in the moment (eg, more effort invested, better attention and focus, or reduced response to distractions), which should improve treatment efficacy without an accompanying increase in exposure.

Current digital mental health interventions require the attention and motivation of patients [5], which are unfortunately also characteristics that are in short supply for people who suffer from depression and could benefit most from treatment (Ibid). Previous work showed that avatar customization can increase motivation to engage with a training application and thus increase the time spent in training. In this paper, we suggest that employing motivational design principles—specifically the use of avatar customization—can increase the efficacy of a Web-based ABMT task by increasing task engagement *in the moment of training*, without requiring additional exposure to the treatment.

Methods

Research Questions

We conducted an online study in which we asked half of the participants to customize an avatar; the other half were assigned a generic avatar. Each avatar group performed an ABMT task—half of the participants in each group were trained to preferentially attend to neutral over negative stimuli, the other half were not. Trained participants should be more resilient to subsequent negative stimuli; thus, participants were all subjected to a negative mood induction (viewing gruesome images) after completing the ABMT task and completed state anxiety scales before training and following the stimulus. Our experiment was a 2 (avatar: customized, generic) by 2 (attentional training, no training) between-subjects design; see Figure 1. Being inattentive, unfocused, or distracted will diminish the efficacy of attention-based training; thus, we ask whether the increased in-the-moment engagement as a result of avatar customization can improve training efficacy.

The following research questions (RQs) guided our analyses:

RQ1: Does customization improve the efficacy of attentional retraining?

RQ2: Does customization increase identification for trained and nontrained participants?

RQ3: Does avatar identification increase the efficacy of attentional training?

RQ4: Does training efficacy vary depending on basic needs satisfaction?

Customization Using an Avatar Creator

To introduce customization, we used an avatar creator that has been shown to facilitate intrinsic motivation and invested effort in a game [20]. Participants were asked to create an avatar, choose its gender, and adjust its appearance, personality, and attributes (characteristics) in the same manner as described in the study by Birk et al [20]; see Figure 2. A minimum of 4 min in the character creator were required, but participants could take longer if they wished. After customizing their avatar, participants were shown a summary of their character.

We presented half of our participants with the avatar creator; the other half were assigned an avatar that had generic features and medium hair and skin color (see Figure 3). Participants were asked their gender in a demographics survey; those who answered male or female were assigned an avatar of the same gender. Those who answered *other* (n=3) were then asked to choose a gender for their digital representation. For both generic and customized avatars, we created and stored 2 facial expressions: 1 neutral face and 1 angry face (see Figure 3). The faces were created by adjusting the 3D model of the face using an algorithm based on Ekman facial action coding system [61]. Due to the differences in facial geometry between male and female faces, there were 2 algorithms (1 for male, 1 for female), and all avatars of the same gender had the same algorithm applied. Once the avatars were customized or assigned, participants completed the ABMT task.

Attention Bias Modification Training Task

In each trial of the ABMT task [31], a fixation cross was presented for 500 ms centered in the screen. Following the presentation of the fixation cross, 2 avatar faces were displayed: 1 with a neutral facial expression and 1 with an angry expression. One face was displayed above the cross and 1 was displayed below the cross. After the faces were presented for 500 ms, they disappeared, and a probe was displayed behind one of the faces. The probe indicated either a right arrow or left arrow, and participants were asked to press the corresponding arrow key on the keyboard as quickly as possible. After pressing the left or right arrow key, an intertrial interval, showing a white screen, was displayed for 500 ms. The next trial started immediately with the presentation of the fixation cross. See Figure 4.

Figure 1. Experimental flow from consent (top) to debriefing. The experimental conditions are highlighted in grey. The negative mood induction is highlighted in orange. BPNS: Basic Psychological Need Satisfaction; ABMT: Attention Bias Modification Training.

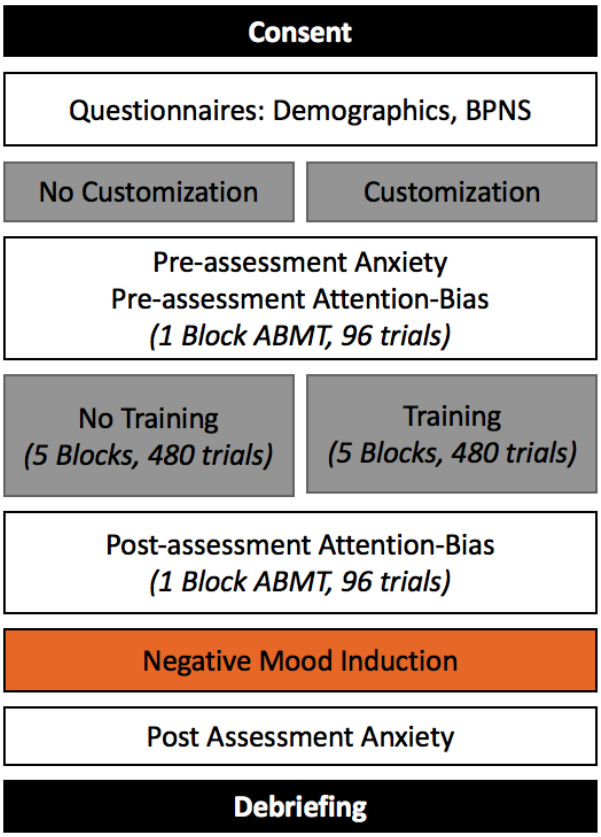


Figure 2. Picture of the avatar creator displaying the customization tool box on the left and a female avatar on the right.

Figure 3. Customized (top) and generic (bottom) male and female avatars with neutral and angry facial expressions.

Figure 4. Attention Bias Modification Training trials. From left to right: fixation cross (500 ms), neutral/angry face (500 ms), probe (participants response), and intertrial interval (500 ms).

Attention Bias Modification Training: Training/No Training

Before beginning the ABMT task, participants were guided through a tutorial of 10 trials of neutral faces to learn the mechanics of the task. Participants were prompted to focus on the fixation cross and could only proceed if they pressed the correct arrow key as indicated by the probe.

Following the tutorial, the ABMT task was presented in 7 blocks: 1 preassessment block, 5 consecutive training/no training blocks, and a postassessment block. During each block, participants completed 96 trials (672 total). After each block, participants had a 6-second break, while being notified that the next block was about to start. This was done to indicate progress through the task.

We used 2 image sets of 4 avatar faces (2 male, 2 female) for the pre- and postassessment blocks; the order of the presentation of the 2 image sets was fully counterbalanced to avoid order effects. The avatars were selected from user-generated avatars from a previous study [20]. The presentation of sex (male,

female), position (top, bottom), probe location (angry, neutral), and probe direction (left, right) was fully balanced over the 96 trials.

In the ABMT training condition, probes only appeared behind neutral faces, with the intent to shift participants' attention toward neutral stimuli. In the no training condition, participants were exposed to a fully balanced probe presentation similar to the assessment condition in which 50.0% (240/480) of the probes appear behind the neutral faces and 50.0% (240/480) behind the angry faces. See the study by Hakamata et al [31] for a detailed description of the ABMT task.

After each of the 7 trail blocks, participants completed a mood scale with 7-point agreement to 4 states (I feel: relaxed, happy, depressed, anxious) representing the 4 corners of arousal-valence space [62].

Trial Removal and Logging

Incorrect responses to the probes were removed from subsequent analysis because they show that the participant did not pay attention in the trial. Individual trials were also removed when participant's response time was greater than 3 SDs of their own mean performance over both assessment blocks. We logged the position (top, bottom), gender (male, female), and expression (angry, neutral) of the probe location; the response time (in ms); and the correctness of the probe response (true, false).

Negative Mood Induction

To measures resilience to a negative mood induction, we presented 20 negative images from the International Affective Picture System (IAPS; [63])—ID: 2703, 3010, 3015, 3225, 3230, 3350, 3530, 3550.1, 9040, 9265, 9301, 9410, 9420, 9433, 9490, 9500, 9570, 9611, 9635.1, and 9901. Images were selected based on valence (mean=2.01, SD=0.5, min=1.51, max=3.60) and arousal (mean=5.92, SD=0.63, min=4.34, max=7.16). To ensure that participants looked at the images, we asked them to rate each image using the valence and arousal scales of the visual self-assessment manikin [64]—valence and arousal scales were sequential to increase the time spent looking at each image. IAPS images have previously been used as a negative mood induction in the context of the ABMT [31]. Descriptively, participants show a similar response to the negative images as with normative IAPS ratings (valance: mean=1.91, SD=0.62, min=1, max=4.05; arousal: mean=4.97, SD=1.215, min=1, max=7), indicating that the images were perceived as expected.

Participants and Deployment Platform

We recruited 317 participants through Amazon's Mechanical Turk (MTurk). MTurk acts as a broker between parties offering a range of human intelligence tasks and paid workers. Although MTurk has been shown to be reliable as a recruitment tool for research [65-67], we excluded participants from analysis if they were not performing the task with care, which we determined in multiple steps. We removed 33 participants based on missing trials (indicating technical difficulties) or too many trials (indicating that they reloaded the task part way through). Then, we calculated variance within each survey subscale and removed participants (n=8) from subsequent analyses who demonstrated response variance greater than 3 SDs above mean variance on 3 or more questionnaire subscales. Having high variance within a subscale is indicative of not paying attention to the survey questions and the reverse-coded items. We also removed participants from subsequent analyses who completed 2 or more questionnaires with an average time below 1 SD of the average response time (n=10). Finally, we removed participants who spent more than 2 min answering the state anxiety questionnaire after the negative mood induction (n=4), as it would indicate that they were taking time to recover while answering the questions.

After the outlier participants were excluded, 262 participants remained. As we controlled our analysis for age and gender, we also excluded the 3 participants who identified their gender as *other*, leaving 259 participants (51.0% [132/259] female, mean age=35.3 years, SD=11.5) in all of our analyses. Participants received compensation of US $10 for their participation. Ethical approval was obtained from the University of Saskatchewan behavioral research ethics board, and participants were asked to give informed consent at the beginning of the task. To comply with ethical guidelines, the task was only available to workers who were older than 18 years. Additionally, only workers from the United States with an approval rate above 90% were offered the task as a means of quality control, and a trigger warning was provided at multiple points before the negative images being presented.

Measures

Identification was measured using the avatar-related subscales of similarity identification, embodied identification, and wishful identification from the Player Identification Scale [68]. Participants rated their agreement to identification-related statements—"My character is like me in many ways"—on a 7-point Likert scale. Identification has been shown before to be an important construct factor when customizing avatars [20].

State anxiety was measured using the state scale from State-Trait Anxiety Inventory (STAI; [69]). Participants rated how well statements—for example, "I'm calm"—described their current state on a 4-point scale from "Not at all" to "Very much." STAI has been successfully been used before in research on resilience to negative affect [70].

Basic needs satisfaction was measured using the Basic Psychological Need Satisfaction (BPNS; [71]) scale. BPNS scale includes subscales for the basic satisfaction of competence, autonomy, and relatedness, as 3 ongoing needs that when satisfied lead to optimal development and function. Participants rated their agreement to statements—"People I know tell me I am good at what I do"—on a 7-point Likert scale. The scale has been used before in research on resilience to negative affect [71].

Procedures

Participants were informed about time (60 min) and payment (US $10), procedure, and the fact that the study included gruesome imagery. After giving consent, participants were asked to fill in questionnaires on their demographics, basic needs satisfaction, and the questionnaire for baseline state anxiety.

Table 1. Descriptive statistics for dependent and control variables displayed by avatar customization and training condition.

Avatar customization	ABMT[a] training	
	No training	Training
STAI[b] (pre), mean (SD)		
No customization	1.95 (0.56)	2.05 (0.64)
Customization	1.94 (0.58)	1.98 (0.57)
STAI (post), mean (SD)		
No customization	2.37 (0.59)	2.6 (0.67)
Customization	2.55 (0.58)	2.37 (0.59)
Identification, mean (SD)		
No customization	2.63 (1.35)	2.62 (1.29)
Customization	4.11 (1.33)	3.76 (1.27)
Relatedness, mean (SD)		
No customization	4.97 (1.29)	4.97 (1.08)
Customization	5.15 (1.02)	5.02 (1.13)
Age (years), mean (SD)		
No customization	37.75 (10.83)	37.75 (10.83)
Customization	32.2 (9.29)	32.2 (9.29)
Gender, n (%)		
Male		
No customization	33 (51.6)	27 (46.6)
Customization	39 (54.9)	34 (51.5)
Female		
No customization	31 (48.4)	31 (53.4)
Customization	32 (45.1)	32 (48.5)

[a]ABMT: Attention Bias Modification Training

[b]STAI: State-Trait Anxiety Inventory

Participants were then assigned to 1 of the 4 conditions: customized avatar/ABMT training, customized avatar/no ABMT training, generic avatar/ABMT training, and generic avatar/no ABMT training. For descriptive statistics, see Table 1.

Depending on the condition, participants either customized an avatar or were assigned a generic avatar and started immediately with the ABMT tutorial. After the tutorial, all participants did a block of preassessment trials with 1 of the previously described image sets. Following this initial block, half of the participants received 5 blocks of attentional retraining, with probes only behind neutral faces, whereas the other half completed 5 blocks, with the probe appearing equally behind neutral and angry faces. All participants performed 1 block of post assessment after training/no training with the other image set. Following assessment, all participants were exposed to the negative mood induction using the gruesome IAPS images. Following the negative mood induction, participants filled in the post state anxiety questionnaire. Finally, participants provided information about how much they identified with the avatar face that was used for training/no-training. Participants were then debriefed about the purpose of the experiment and were directed to adorable images of baby animals if they wanted to combat the negative mood induced by the IAPS images. See Figure 1 for the experimental flow.

Data Collection and Analysis

All data were logged to a database on a server at the University of Saskatchewan and were analyzed using Statistical Package for the Social Sciences (SPSS) version 24 (IBM Corp) with the process macro for moderation and mediation analyses [72].

Before analyzing the effects of training and customization on anxiety, we analyzed the block-based mood ratings to ensure that the different conditions did not directly induce different vulnerability to the negative mood induction. There were no significant interactions with either avatar customization or training, implying that the conditions did not directly influence mood (see [73]).

We computed 4 models. First, we investigated the role of avatar customization and attentional retraining on state anxiety after the negative mood induction. We use an analysis of covariance (ANCOVA) with avatar (customized, generic) and attentional retraining (training, no training) as factors on the dependent

measure of state anxiety measured after the negative mood induction.

The ANCOVA allowed us to control for levels of state anxiety before attentional retraining, age, and sex by including these variables as covariates. ANCOVAs have been shown to have higher power in randomized studies with a baseline compared with using repeated measures analysis of variances (ANOVAs) [74].

The next 3 regression models were all controlled for baseline state anxiety, age, and gender. In our second model, we considered whether avatar customization (X) predicts identification (Y) and whether training (M) moderates this relationship (model=1; [72]). The link between avatar customization and identification has been previously established [20,75].

Third, we considered whether identification (X) predicts anxiety (Y; post negative mood induction) and whether training (M) moderates this relationship (model=1; [72]).

Finally, we conducted a moderated moderation (model=3; [72]). We considered whether avatar customization (X) predicts anxiety post negative mood induction (Y), whether this relationship is moderated by baseline need satisfaction (M), and whether that moderation is moderated by training (W). A moderated moderation model is similar to a 3-way interaction in an ANOVA but allows for the inclusion of continuous variables as factors.

Results

We created 4 statistical models, as described above, and used these models to answer a series of research questions.

RQ1: Does Customization Improve the Efficacy of Attentional Retraining?

Participants in the ABMT training group who customized their avatar should show increased resilience to negative stimuli. An ANCOVA controlling for pretraining anxiety, age, and gender showed a significant interaction between avatar customization and attentional retraining on anxiety measured after the negative mood induction ($F_{1,252}$=6.86, P=.009, η_p^2=.027). Bonferroni-corrected post hoc tests showed that training reduces state anxiety only when participants used a customized avatar (P=.04); when a generic avatar was used, training was not more effective than no training (P=.10). This result suggests that attentional retraining is more effective when participants were allowed to create a customized avatar and were presented with personalized stimuli (see Figure 5).

Although the interaction of avatar and training shows a significant effect, the partial eta-squared value (η_p^2=.027) implies a small effect. The large effect on poststimulus anxiety (R^2=.401) is mostly explained by baseline anxiety (η_p^2=.353). However, when looking at the model by individual effect sizes of the included variables (age, gender, training, avatar, and the interaction of training and avatar), we find that age (η_p^2=.032) and gender (η_p^2=.030) both show small effects. Training and avatar (both not significant) explain almost no variance (η_p^2<.001 and η_p^2=.001, respectively). The interaction of avatar and training (η_p^2=.027) shows an effect similar in size to the effects of major demographic variables (ie, age and gender) on anxiety.

These results are in line with previous research, showing that customization increased identification with a digital representation, and as a result increased positive affect, task enjoyment, invested effort, and motivated behavior measured as time spent in a free choice game [20], and that these effects are of a similar size as the effects of major demographic variables.

RQ2. Does Customization Increase Identification for Trained and Nontrained Participants?

We investigated the prediction of avatar customization on avatar identification, moderated by training. The model was significant ($F_{5,252}$=12.46, P<.001, R^2=.23). Customization predicts identification (beta=.651, P<.001); however, the nonsignificant interaction with training ($F_{1,252}$=0.79, P=.38) shows that the prediction does not depend on whether the probe appeared under angry or both angry and neutral faces (see Figure 6).

RQ3. Does Avatar Identification Increase the Efficacy of Attentional Training?

We investigated the prediction of avatar identification on anxiety measured after the negative mood induction, as moderated by training. The model was significant ($F_{6,252}$=28.49, P<.001, R^2=.40). There was neither a significant main effect for identification (P=.99) nor for training (P=.81). However, the interaction between identification and training ($F_{1,252}$=6.19, P=.01) shows that for participants who underwent training, identification tends to reduce anxiety, that is, increases training efficacy; however, for participants who did not undergo training, the trend is in the other direction—greater identification increases anxiety (see Figure 5).

RQ4. Does Training Efficacy Vary Depending on Basic Needs Satisfaction?

We showed in R1 that avatar customization and training interact to yield lower anxiety. Now we ask whether this effect is being driven by participants with varying levels of basic psychological needs satisfaction. We used basic satisfaction of relatedness as it has been shown to be an important predictor of depression, addiction, and other mental disorders [76]. A moderation model with the predictor avatar (X) on state anxiety (Y) moderated by satisfaction of relatedness (M) and then moderated by training (W) was significant ($F_{10,248}$=18.5, P<.001, R^2=.43); we controlled for baseline state anxiety, gender, and age. As expected, the interaction between avatar and training was significant (P=.007). Most interestingly, the interaction between satisfaction of relatedness, avatar, and training also was significant (P=.02). This 3-way interaction showed that the interaction between avatar customization and training is more pronounced for those lower in satisfaction of relatedness (see Figure 7).

Figure 5. Left: state anxiety for the interaction of training and avatar customization (RQ1). Right: state anxiety for the interaction of training and identification (RQ3). Displayed data are controlled for age, gender, and baseline state anxiety.

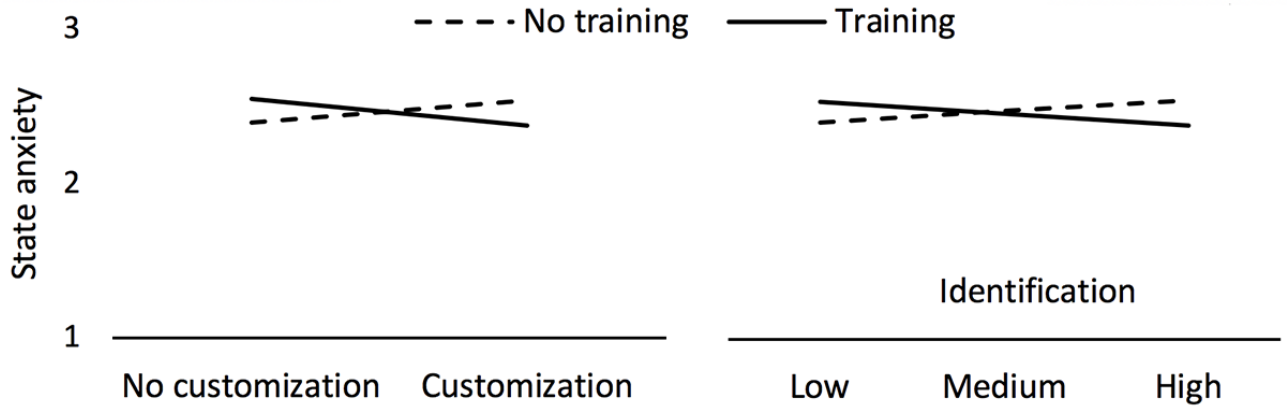


Figure 6. Effect of training and customization on identification (RQ2). Displayed data are controlled for age and gender.

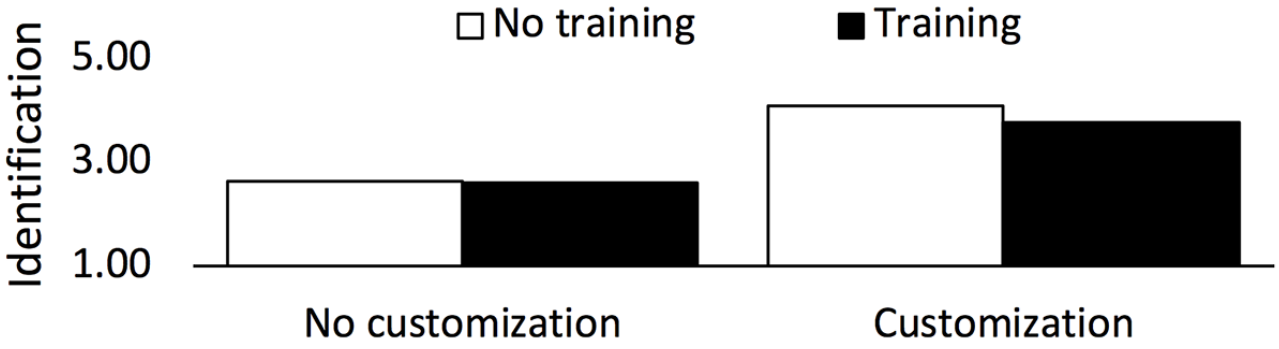
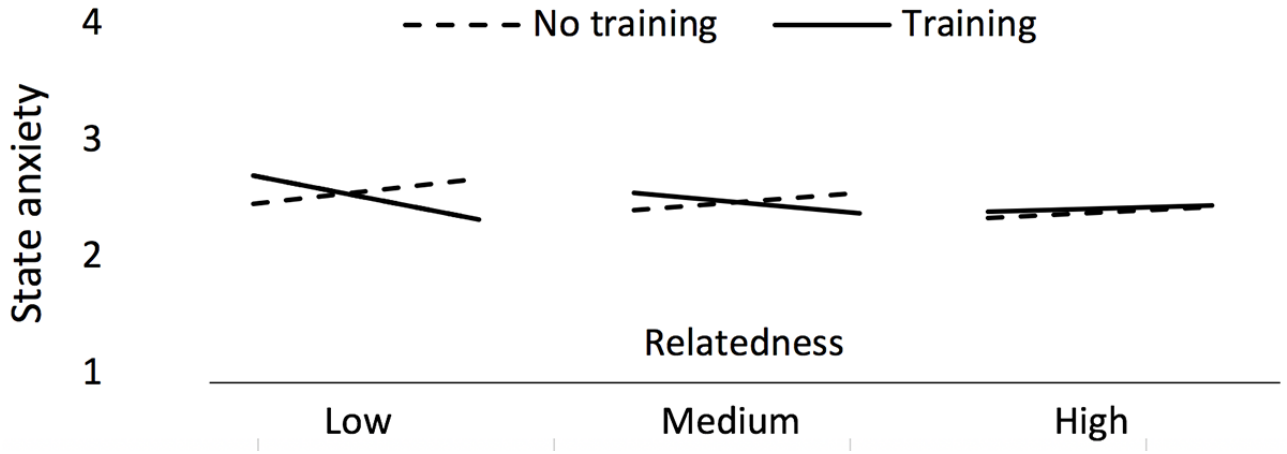


Figure 7. Moderation model showing avatar on state anxiety moderated by satisfaction of relatedness and then moderated by training (RQ4). Displayed data are controlled for age, gender, and baseline state anxiety.



Participants with lower relatedness satisfaction were driving the significant interaction between avatar customization and training, which is a meaningful result, as these are the participants who are most likely to be in need of a mental health intervention in the first place [76].

Discussion

We summarize our findings, contextualize them in theory, and discuss the implications, limitations, and opportunities for future work.

Principal Findings

The results revealed 4 main findings:

1. First, avatar customization increased resilience to a negative mood induction for participants who trained attentional bias, compared with those who did not train attentional bias. Thus, avatar customization improved training efficacy, presumably as a result of increased in-the-moment engagement.
2. Second, avatar customization increased avatar identification, regardless of whether or not participants trained attentional bias.
3. Third, avatar identification tended to reduce anxiety after the negative mood induction for participants who underwent training but tended to increase poststimulus anxiety for

participants who did not undergo training, suggesting that customization increases in-the-moment engagement.

4. Fourth, the beneficial effect of avatar customization on training is being driven by participants who are low in their basic satisfaction of relatedness, which is important because these are the participants who are most likely in need of digital interventions for mental health.

Explanation of Findings

Our findings suggest that attentional retraining is improved when participants are more engaged and invested in the task. On the basis of prior research [20], we assume that increased motivation and willingness to invest effort increases attention to the task and efficacy as a result. We further explain potential mechanisms of why avatar customization increased treatment efficacy.

Self-Determined Experience

The results presented in this study and previous research using avatar customization [20] show that customization increases identification and that identification positively predicts task engagement. Engagement can be understood as an increase in participants' invested effort, a central construct in describing intrinsic motivation and task engagement [77]. Research on motivation [78,79] suggests that invested effort depends on factors related to the experience itself, in particular, whether or not the interaction satisfies our needs for competence (ie, feeling a sense of mastery over challenges), autonomy (ie, experiencing the choice to engage under our own volition), and relatedness (ie, feeling connected to others) [80]. In our experiment, we aimed to increase motivation by manipulating choice as a means of increasing autonomy, that is, participants either customized an avatar or they were assigned a generic avatar. Exercising choice has motivational benefits, increasing the sense of volitional control and ownership over analog and digital objects [81,82].

SDT further suggests that intentionally engaging in a task is an important predictor of outcome— presumably because of the increased attention and focus paid to the task at hand. A more positive attitude toward the task—facilitated, for example, by feeling volitional control over the content or ownership over objects—may lower cognitive resistance to engagement. Participants in the customization condition may have been less likely to reject engaging with the task simply because it did not appeal to their preferences. The ABMT—like other CCT tasks—is not a particularly entertaining task, so frustration and boredom have the potential to undermine the positive effects of the training; however, we assume potential boredom and frustration are partially counteracted by the positive effects of customization.

Personalized Experience

We have argued that avatar customization increased efficacy by increasing autonomy and invested effort or by improving attitude and feelings of control. In addition, the customized avatar condition presents tailored stimuli, that is, customized avatar faces with neutral or negative expressions, during the training phase. Tailoring (through customization or personalization) is a persuasive strategy that has been used, for example, to increase motivation to play serious games [83] and to improve the efficacy of serious games for changing attitudes, intentions, and self-efficacy around healthy eating behavior [84]. The personalization of the stimuli in our experiment may have directly affected participants' perception of the stimuli, making the stimuli's emotions (neutral, negative) more salient to them.

Individual Differences

Our results suggest that the customized avatar increases engagement in the moment of use and subsequently training efficacy; however, we also show that the improvements resulting from avatar customization were most pronounced for those participants who experience low satisfaction of relatedness, that is, people who feel less connected to others, experience less support, or even feel lonely. Our intervention facilitates choice but enhances the experience of relatedness in the application. Although we did not explicitly measure relatedness satisfaction during the task, prior research has shown that customization can facilitate feelings of connectedness with a digital representation [80]. Increased satisfaction of relatedness might have improved engagement with the training task involving an avatar. This is important because loneliness, feelings of social exclusion, and feeling disconnected from supportive social groups are predictors of vulnerability for many mental health conditions such as depression and anxiety [85].

Increased Engagement With No Training

Avatar customization increased resilience to negative stimuli after the attentional bias modification training (ie, saw probes appear only under neutral faces); however, we also observed that the group that used avatar customization without training (ie, saw probes appear under both angry and neutral faces) was actually *more* susceptible to the negative mood induction. These results suggest that avatar customization increased task engagement, independent of the training condition. For participants who trained, this resulted in better resilience to the negative stimuli than those who trained with a generic avatar face; however, for participants who did not train, this likely resulted in them investing more attention overall and this being more susceptible to the negative mood induction than those who did not train using the generic avatar faces. This interaction result is not surprising, considering prior evidence of the relationship between avatar customization, identification, and motivation [20] in increasing task engagement, and further supports our arguments for customization as a motivational design strategy that can increase in-the-moment engagement.

Blending and Extending Existing Therapeutic Approaches

There are a variety of existing approaches to maximize the efficacy of internet-based interventions in mental health. For example, blended interventions, that is, interventions that blend internet-based forms of therapy with the interleaved presence of a therapist, have been shown to decrease the load on therapists and are similar in effect to traditionally delivered cognitive-behavioral therapy (CBT) programs [86,87].

Stand-alone internet-based CBT programs [88,89] have also been evaluated in RCTs and have shown promise for the

improvement of mental health and well-being. For example, modular internet-based CBT programs such as MoodGYM [6], SilverCloud [90], or the mobile-based suite IntelliCare [3] have been shown to successfully improve mental health in cases of depression or anxiety. However, limitations of internet-based CBT programs such as the reading and language requirements, financial and time costs of localization, and the required time investment that is difficult for users to achieve suggest that other approaches might need to be deployed in tandem. Furthermore, CCTs such as ABMT have been applied adjacent [91] to internet-based CBTs, showing potential in blending approaches.

Although our study only considers customization in the context of bias modification, there are other approaches to internet-based intervention design that may benefit from the increased task engagement that we demonstrate. For example, including customized avatars to guide a patient through a CBT application or including other personalized stimuli could potentially increase engagement in the moment or result in increased adherence, as suggested by Birk et al [21]. Or consider digital phobia treatments that expose patients to fear-inducing stimuli (eg, [92-94]), which could be even more effective if patients are able to customize the presented stimuli, thereby increasing salience in an individual patient's personal context. And finally, consider narrative-based therapeutic applications for people who suffer from post-traumatic stress disorder, which walk a patient through their experiences and help them to reframe the traumatic event (eg, [95,96]). Supporting patients to personalize the narrative, graphical objects, and other intervention elements could increase the efficacy of this important type of mental health intervention.

Until we validate innovative methods, adapting well-evaluated approaches to be delivered at scale is a safe and promising way forward. Exposing people to techniques that are not ready for use as treatment has associated dangers, whereas investigating subtle adaptations to existing approaches—such as interface customization applied to existing treatments—may improve and optimize established interventions.

Interface customization is not an intervention in itself but rather an enhancement that can be applied across a range of existing interventions. From the perspective of implementation medicine—the branch of medicine that asks if a research result should be implemented in practice [97]—it is a requirement to have evidence-based proof that a technique is either more effective than prior techniques or at least provides equal effectiveness with a lower investment of time and/or money [98]. Our research shows that tweaks to existing interventions such as adding customization can significantly improve in-the-moment engagement.

Supporting Motivation in the Immediate and Long Term

In this research, we show how avatar customization can increase task engagement and focused attention in the moment of application use. Motivational theories—such as SDT—describe how fostered autonomy can increase the effort invested in a task, which we suggest results in subsequent task engagement during intervention use. In contrast, our previous work on avatar customization has shown increased adherence to a daily training regimen over a medium-term (3-week) breathing intervention [21]. It is important to distinguish the motivational benefits that accrue from increased exposure to treatment (ie, through increased log-ins or more persistent usage) and those that result from increased attention in the moment of intervention use, without any accompanying increase in treatment exposure. Customizing avatars has been shown to increase motivation both in terms of increased treatment exposure through retention [21] and—in this paper—through focused attention in the moment. Additionally, personalization has been explored as an approach to enhance long-term engagement in the context of digital interventions for mental health [60]. Further work is needed to explore the potential additive effects of these 2 approaches in increasing motivation and to explore the mediating motivational factors (eg, attention, enjoyment, and effort) that could explain the improved outcome observed in this study.

Limitations and Future Work

Our study has limitations that we intend to address in future work.

First, identification occurs in multiple ways, that is, wishful, similarity, and embodied identification [99]; in this paper, we did not manipulate the different aspects of identification, which potentially could enhance or diminish the efficacy of customization. Using, for example, nonhumanoid representations in a customization procedure might differentially affect identification and efficacy of a subsequent task. We plan to investigate different types of customization to explore the differential effects of identification types on intervention efficacy and to determine whether customization of other interface elements (beyond faces) can have the same benefits.

Second, the experimental context needs to be considered for the interpretation of our results; participants were recruited using a crowdsourcing platform and were paid for their participation, which creates a different experience than being exposed to attentional retraining as part of a therapy for participants who are in need of treatment.

Third, the attention bias modification task is only 1 potential digital intervention. Enhancing efficacy through customization needs to be tested across various interventions. Enhanced engagement through customization will particularly benefit interventions that rely on focused attention, such as cognitive tasks like the ABMT. How customization can integrate with other therapeutic approaches (eg, CBT-based interventions) remains to be investigated. Our technique requires very little effort to implement but shows significant changes of efficacy. How to leverage volitional engagement to best increase the efficacy of a variety of interventions is a promising direction; however, more systematic research on the limits of customization for in-the-moment task engagement needs is required.

Fourth, our study does not distinguish whether the effects are a result of participants coping better with the negative mood induction post training or if the effects are a result of the training itself. The former would suggest that the participants developed better strategies to disconnect from maladaptive thought processes, whereas the latter suggests that training helps

participants protectively shift attention away from negative cognitions.

Fifth, this study focuses on the effects of in-the-moment engagement. However, many tasks (including ABMT) require frequent repetitions, for example, being used for 30-min daily. Although previous work suggests that avatar customization has positive effects on long-term engagement [21], we did not specifically investigate the effects of avatar customization over the long term in this study.

Conclusions

Mental illness is increasing, yet therapies have not adapted to meet the growing demand. Computerized interventions delivered at scale have the potential to ease clinical demand and interventions accessible for those who do not qualify for traditional therapies or cannot access or afford them. However, the efficacy of computerized training cannot be sacrificed in service of a wider reach. Our results suggest that increasing in-the-moment engagement through interface customization and personalization can increase training efficacy.

We asked participants to complete online ABMT with a customized avatar or an assigned generic avatar. ABMT helps people shift their attention away from negative stimuli and has been shown to increase affective resilience to a subsequent negative mood induction (eg, rating gruesome images). Our results showed that a version of the ABMT using customized interface elements generated through avatar customization increased resilience to a subsequent negative mood induction, suggesting that avatar customization increases in-the-moment engagement, and subsequently training efficacy. Furthermore, the customization benefits were particularly pronounced for participants with low satisfaction of relatedness, who are most at risk for developing mental illness.

Digital interventions delivered at scale offer a promise of increasing the reach of mental health treatment to a greater number of people in a wider range of places. Our work shows that avatar customization may help to improve the efficacy of existing and future training programs delivered in the wild.

Acknowledgments

We thank Natural Sciences and Engineering Research Council for funding, Benj Hingston and Cher Atkins for their work on the attentional bias modification training and the avatar customizer, Jason Bowey and Colby Johanson for implementing the Bride of the Franken-System, all Mechanical Turk workers for working with us, and David Guetta for providing guilty pleasure when needed.

Conflicts of Interest

None declared.

References

1. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
2. Mandryk RL, Birk MV. Toward game-based digital mental health interventions: player habits and preferences. J Med Internet Res 2017 Dec 20;19(4):e128 [FREE Full text] [doi: 10.2196/jmir.6906] [Medline: 28428161]
3. 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 Dec 05;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
4. 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]
5. 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 1;68(5):427-429. [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
6. Christensen H, Griffiths KM, Jorm AF. Delivering interventions for depression by using the internet: randomised controlled trial. BMJ 2004 Jan 31;328(7434):265 [FREE Full text] [doi: 10.1136/bmj.37945.566632.EE] [Medline: 14742346]
7. Thiart H, Lehr D, Ebert DD, Berking M, Riper H. Log in and breathe out: internet-based recovery training for sleepless employees with work-related strain - results of a randomized controlled trial. Scand J Work Environ Health 2015 Mar;41(2):164-174 [FREE Full text] [doi: 10.5271/sjweh.3478] [Medline: 25590336]
8. Meyer B, Bierbrodt J, Schröder J, Berger T, Beevers CG, Weiss M, et al. Effects of an Internet intervention (Deprexis) on severe depression symptoms: randomized controlled trial. Internet Interv 2015 Mar;2(1):48-59. [doi: 10.1016/j.invent.2014.12.003]
9. 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] [doi: 10.1136/bmj.h5627] [Medline: 26559241]

10. Mohr DC, Ho J, Duffecy J, Reifler D, Sokol L, Burns MN, et al. Effect of telephone-administered vs face-to-face cognitive behavioral therapy on adherence to therapy and depression outcomes among primary care patients: a randomized trial. JAMA 2012 Jun 6;307(21):2278-2285 [FREE Full text] [doi: 10.1001/jama.2012.5588] [Medline: 22706833]
11. Ryan RM, Patrick H, Deci EL, Williams GC. Facilitating health behaviour change and its maintenance: interventions based on self-determination theory. Health Psychol Rev 2008 Mar;10(1):2-5.
12. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
13. Deterding S, Dixon D, Kahled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. New York: ACM; 2011 Sep 28 Presented at: Academic MindTrek; September 28 - 30, 2011; Tampere p. 9-15. [doi: 10.1145/2181037.2181040]
14. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychol Bull 1999 Nov;125(6):627-68; discussion 692. [Medline: 10589297]
15. Armitage CJ, Conner M. Efficacy of the Theory of Planned Behaviour: a meta-analytic review. Br J Soc Psychol 2001 Dec;40(Pt 4):471-499. [Medline: 11795063]
16. Mandryk RL, Dielschneider S, Kalyn MR, Bertram CP, Gaetz M, Doucette A, et al. Games as neurofeedback training for children with FASD. In: Proceedings of the 12th International Conference on Interaction Design and Children. 2013 Presented at: Interaction Design and Children; June 24-27, 2013; New York p. 165-172. [doi: 10.1145/2485760.2485762]
17. Smeddinck JD, Herrlich M, Malaka R. Exergames for Physiotherapy and Rehabilitation: A Medium-term Situated Study of Motivational Aspects and Impact on Functional Reach. In: Exergames for Physiotherapy and Rehabilitation.: ACM Press; 2015 Apr 18 Presented at: Human Factors in Computing Systems; April 18 - 23, 2015; Seoul, Republic of Korea p. 4143-4146. [doi: 10.1145/2702123.2702598]
18. Hernandez HA, Ketcheson M, Schneider A, Ye Z, Fehlings D, Switzer L, et al. Design and evaluation of a networked game to support social connection of youth with cerebral palsy. In: Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility. New York: ACM; 2014 Oct 20 Presented at: ACM SIGACCESS conference on Computers & accessibility; October 20 - 22, 2014; Rochester, New York p. 161-168. [doi: 10.1145/2661334.2661370]
19. Birk MV, Mandryk RL, Atkins C. The motivational push of games: The interplay of intrinsic motivation and external rewards in games for training. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. New York: ACM; 2016 Oct 16 Presented at: Annual Symposium on Computer-Human Interaction in Play; October 16 - 19, 2016; Austin, Texas, USA p. 291-303. [doi: 10.1145/2967934.2968091]
20. Birk MV, Atkins C, Bowey JT, Mandryk RL. Fostering Intrinsic Motivation through Avatar Identification in Digital Games. Proc CHI Conf Hum Factors Comput Syst - CHI ?16 .?2995. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. New York: ACM; 2016 May 7 Presented at: Conference on Human Factors in Computing Systems; May 07 - 12, 2016; San Jose, California, USA p. 2982-2995. [doi: 10.1145/2858036.2858062]
21. Birk MV, Mandryk RL. Combating Attrition in Digital Self-Improvement Programs using Avatar Customization. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM Press; 2018 Apr 21 Presented at: Conference on Human Factors in Computing Systems; April 21 - 26, 2018; Montreal QC, Canada p. 660. [doi: 10.1145/3173574.3174234]
22. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med 2014 Nov;11(11):e1001756 [FREE Full text] [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
23. Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. The Lancet 2015 Jun 6;385(9984):2255-2263. [doi: 10.1016/S0140-6736(15)60461-5] [Medline: 25771249]
24. Motter JN, Pimontel MA, Rindskopf D, Devanand DP, Doraiswamy PM, Sneed JR. Computerized cognitive training and functional recovery in major depressive disorder: a meta-analysis. J Affect Disord 2016 Jan 1;189:184-191. [doi: 10.1016/j.jad.2015.09.022] [Medline: 26437233]
25. Motter JN, Devanand DP, Doraiswamy PM, Sneed JR. Computerized cognitive training for major depressive disorder: What's next? Front Psychiatry 2015;6:137 [FREE Full text] [doi: 10.3389/fpsyt.2015.00137] [Medline: 26483706]
26. Coyle H, Traynor V, Solowij N. Computerized and virtual reality cognitive training for individuals at high risk of cognitive decline: systematic review of the literature. Am J Geriatr Psychiatry 2015 Apr;23(4):335-359. [doi: 10.1016/j.jagp.2014.04.009] [Medline: 24998488]
27. Cortese S, Ferrin M, Brandeis D, Buitelaar J, Daley D, Dittmann RW, European ADHD Guidelines Group (EAGG). Cognitive training for attention-deficit/hyperactivity disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. J Am Acad Child Adolesc Psychiatry 2015 Mar;54(3):164-174 [FREE Full text] [doi: 10.1016/j.jaac.2014.12.010] [Medline: 25721181]

28. Bogdanova Y, Yee MK, Ho VT, Cicerone KD. Computerized cognitive rehabilitation of attention and executive function in acquired brain injury: a systematic review. J Head Trauma Rehabil 2016;31(6):419-433 [FREE Full text] [doi: 10.1097/HTR.0000000000000203] [Medline: 26709580]
29. Gomez P, Ratcliff R, Perea M. A model of the go/no-go task. J Exp Psychol Gen 2007 Aug;136(3):389-413 [FREE Full text] [doi: 10.1037/0096-3445.136.3.389] [Medline: 17696690]
30. Klingberg T, Fernell E, Olesen PJ, Johnson M, Gustafsson P, Dahlström K, et al. Computerized training of working memory in children with ADHD--a randomized, controlled trial. J Am Acad Child Adolesc Psychiatry 2005 Feb;44(2):177-186. [doi: 10.1097/00004583-200502000-00010] [Medline: 15689731]
31. Hakamata Y, Lissek S, Bar-Haim Y, Britton J, Fox NA, Leibenluft E, et al. Attention bias modification treatment: a meta-analysis toward the establishment of novel treatment for anxiety. Biol Psychiatry 2010 Dec 1;68(11):982-990 [FREE Full text] [doi: 10.1016/j.biopsych.2010.07.021] [Medline: 20887977]
32. Thorell LB, Lindqvist S, Bergman Nutley S, Bohlin G, Klingberg T. Training and transfer effects of executive functions in preschool children. Dev Sci 2009 Jan;12(1):106-113. [doi: 10.1111/j.1467-7687.2008.00745.x] [Medline: 19120418]
33. Prins PJ, Dovis S, Ponsioen A, ten Brink E, van der Oord S. Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD? Cyberpsychol Behav Soc Netw 2011 Mar;14(3):115-122. [doi: 10.1089/cyber.2009.0206] [Medline: 20649448]
34. van Koningsbruggen GM, Veling H, Stroebe W, Aarts H. Comparing two psychological interventions in reducing impulsive processes of eating behaviour: effects on self-selected portion size. Br J Health Psychol 2014 Nov;19(4):767-782. [doi: 10.1111/bjhp.12075] [Medline: 24147757]
35. Hakamata Y, Lissek S, Bar-Haim Y, Britton JC, Fox NA, Leibenluft E, et al. Attention bias modification treatment: a meta-analysis toward the establishment of novel treatment for anxiety. Biol Psychiatry 2010 Dec 1;68(11):982-990 [FREE Full text] [doi: 10.1016/j.biopsych.2010.07.021] [Medline: 20887977]
36. Jones EB, Sharpe L. Cognitive bias modification: a review of meta-analyses. J Affect Disord 2017 Dec 1;223:175-183. [doi: 10.1016/j.jad.2017.07.034] [Medline: 28759865]
37. Barry TJ, Vervliet B, Hermans D. An integrative review of attention biases and their contribution to treatment for anxiety disorders. Front Psychol 2015;6:968 [FREE Full text] [doi: 10.3389/fpsyg.2015.00968] [Medline: 26217284]
38. Carlbring P, Apelstrand M, Sehlin H, Amir N, Rousseau A, Hofmann SG, et al. Internet-delivered attention bias modification training in individuals with social anxiety disorder--a double blind randomized controlled trial. BMC Psychiatry 2012 Jun 25;12:66 [FREE Full text] [doi: 10.1186/1471-244X-12-66] [Medline: 22731889]
39. Glenn D, Golinelli D, Rose RD, Roy-Byrne P, Stein MB, Sullivan G, et al. Who gets the most out of cognitive behavioral therapy for anxiety disorders? The role of treatment dose and patient engagement. J Consult Clin Psychol 2013 Aug;81(4):639-649 [FREE Full text] [doi: 10.1037/a0033403] [Medline: 23750465]
40. Chu BC, Kendall PC. Positive association of child involvement and treatment outcome within a manual-based cognitive-behavioral treatment for children with anxiety. J Consult Clin Psychol 2004 Oct;72(5):821-829. [doi: 10.1037/0022-006X.72.5.821] [Medline: 15482040]
41. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
42. Joosten EA, DeFuentes-Merillas L, de Weert GH, Sensky T, van der Staak CP, de Jong CA. Systematic review of the effects of shared decision-making on patient satisfaction, treatment adherence and health status. Psychother Psychosom 2008;77(4):219-226. [doi: 10.1159/000126073] [Medline: 18418028]
43. Connolly T, Boyle E, MacArthur E, Hainey T, Boyle J. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ 2012 Sep;59(2):661-686. [doi: 10.1016/j.compedu.2012.03.004]
44. Basak C, Boot WR, Voss MW, Kramer AF. Can training in a real-time strategy video game attenuate cognitive decline in older adults? Psychol Aging 2008 Dec;23(4):765-777 [FREE Full text] [doi: 10.1037/a0013494] [Medline: 19140648]
45. Subrahmanyam K, Greenfield PM. Effect of video game practice on spatial skills in girls and boys. J Appl Dev Psychol 1994 Jan;15(1):13-32. [doi: 10.1016/0193-3973(94)90004-3]
46. Boot WR, Kramer AF, Simons DJ, Fabiani M, Gratton G. The effects of video game playing on attention, memory, and executive control. Acta Psychol (Amst) 2008 Nov;129(3):387-398. [doi: 10.1016/j.actpsy.2008.09.005] [Medline: 18929349]
47. Dorval M, Pépin M. Effect of playing a video game on a measure of spatial visualization. Percept Mot Skills 1986 Feb;62(1):159-162. [doi: 10.2466/pms.1986.62.1.159] [Medline: 3960656]
48. Clark JE, Lanphear AK, Riddick CC. The effects of videogame playing on the response selection processing of elderly adults. J Gerontol 1987 Jan;42(1):82-85. [Medline: 3794204]
49. Green CS, Bavelier D. Action video game modifies visual selective attention. Nature 2003 May 29;423(6939):534-537. [doi: 10.1038/nature01647] [Medline: 12774121]
50. Drew B, Waters J. Video games: utilization of a novel strategy to improve perceptual motor skills and cognitive functioning in the non-institutionalized elderly. Cogn Rehabil 1986;4:26-31.
51. Dustman RE, Emmerson RY, Steinhaus LA, Shearer DE, Dustman TJ. The effects of videogame playing on neuropsychological performance of elderly individuals. J Gerontol 1992 May 1;47(3):P168-P171. [doi: 10.1093/geronj/47.3.P168]

52. Orosy-Fildes C, Allan RW. Psychology of computer use: XII. Videogame play: human reaction time to visual stimuli. Percept Mot Skills 2016 Aug 31;69(1):243-247. [doi: 10.2466/pms.1989.69.1.243]
53. Dye MW, Green CS, Bavelier D. The development of attention skills in action video game players. Neuropsychologia 2009 Jul;47(8-9):1780-1789 [FREE Full text] [doi: 10.1016/j.neuropsychologia.2009.02.002] [Medline: 19428410]
54. Goldstein J, Cajko L, Oosterbroek M, Michielsen M, Van Houten O, Salverda F. Video games and the elderly. Soc Behav Pers 1997 Nov;25:345-352. [doi: 10.2224/sbp.1997.25.4.345]
55. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc; Dec 1, 2002:2.
56. Kaptein M, Markopoulos P, de Ruyter B, Aarts E. Personalizing persuasive technologies: explicit and implicit personalization using persuasion profiles. Int J Hum Comput Stud 2015 May;77:38-51. [doi: 10.1016/j.ijhcs.2015.01.004]
57. Oinas-Kukkonen H, Harjumaa M. A Systematic Framework for Designing and Evaluating Systems. In: Proceedings of the 3rd international conference on Persuasive Technology. 2008 Jun 4 Presented at: International conference on Persuasive Technology; 2008; Oulu, Finland p. 164-176. [doi: 10.1007/978-3-540-68504-3]
58. Orji R, Vassileva J, Mandryk RL. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. User Model User-Adap Inter 2014 Jul 14;24(5):453-498. [doi: 10.1007/s11257-014-9149-8]
59. Sundar SS, Marathe SS. Personalization versus customization: the importance of agency, privacy, and power usage. Hum Commun Res 2010 Jun 7;36(3):298-322.
60. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, USA: ACM; 2012 May 5 Presented at: Conference on Human Factors in Computing Systems; USA — May 05 - 10, 2012; Austin, Texas, USA p. 1421-1430. [doi: 10.1145/2207676.2208602]
61. Ekman P, Friesen WV. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto: Consulting Psychologists Press; 1978.
62. Mandryk RL, Atkins MS. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. Int J Hum Comput Stud 2007 Apr;65(4):329-347. [doi: 10.1016/j.ijhcs.2006.11.011]
63. Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Gainesville, Florida, USA: Technical Report A-6, The Center for Research in Psychophysiology, University of Florida; 2008.
64. Bradley MM, Lang PJ. Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 1994 Mar;25(1):49-59. [Medline: 7962581]
65. Shapiro DN, Chandler J, Mueller PA. Using mechanical turk to study clinical populations. Clin Psychol Sci 2013 Jan 31;1(2):213-220. [doi: 10.1177/2167702612469015]
66. Crump MJ, McDonnell JV, Gureckis TM. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One 2013;8(3):e57410 [FREE Full text] [doi: 10.1371/journal.pone.0057410] [Medline: 23516406]
67. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. Elife 2016 Mar 1;5:e11305 [FREE Full text] [doi: 10.7554/eLife.11305] [Medline: 26928075]
68. Van Looy J, Courtois C, De Vocht M, De Marez L. Player identification in online games: validation of a scale for measuring identification in MMOGs. Media Psychol 2012 May 18;15(2):197-221. [doi: 10.1080/15213269.2012.674917]
69. Spielberger CD. Test anxiety inventory. Hoboken, New Jersey, United States: John Wiley & Sons, Inc; 2010.
70. Davidson RJ, Kabat-Zinn J, Schumacher J, Rosenkranz M, Muller D, Santorelli SF, et al. Alterations in brain and immune function produced by mindfulness meditation. Psychosom Med 2003;65(4):564-570. [Medline: 12883106]
71. Gagne M. Autonomy support and need satisfaction in the motivation and well-being of gymnasts. J Appl Sport Psychol 2003 Oct;15(4):372-390. [doi: 10.1080/714044203]
72. Hayes A. PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, and Conditional Process Modeling [White Paper]. 2012. URL: http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/SobelTest?action=AttachFile&do=get&target=process.pdf [accessed 2018-11-05]
73. MacLeod C, Rutherford E, Campbell L, Ebsworthy G, Holker L. Selective attention and emotional vulnerability: assessing the causal basis of their association through the experimental manipulation of attentional bias. J Abnorm Psychol 2002 Feb;111(1):107-123. [Medline: 11866165]
74. Van Breukelen GJ. ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]. J Clin Epidemiol 2006 Sep;59(9):920-925. [doi: 10.1016/j.jclinepi.2006.02.007] [Medline: 16895814]
75. Trepte S, Reinecke L. Avatar creation and video game enjoyment. J Media Psychol 2010 Jan;22(4):171-184. [doi: 10.1027/1864-1105/a000022]
76. Blatt SJ, Zuroff DC. Interpersonal relatedness and self-definition: two prototypes for depression. Clin Psychol Rev 1992 Jan;12(5):527-562. [doi: 10.1016/0272-7358(92)90070-O]
77. McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exerc Sport 1989 Mar;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]

78. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
79. Gagné M, Deci E. Self-determination theory and work motivation. J Organ Behav 2005 Apr 14;26(4):331-362. [doi: 10.1002/job.322]
80. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
81. Huang Y, Wang L, Shi J. When do objects become more attractive? The individual and interactive effects of choice and ownership on object evaluation. Pers Soc Psychol Bull 2009 Jun;35(6):713-722. [doi: 10.1177/0146167209333046] [Medline: 19329495]
82. Beggan JK. On the social nature of nonsocial perception: The mere ownership effect. J Pers Soc Psychol 1992;62(2):229-237. [doi: 10.1037/0022-3514.62.2.229]
83. Göbel S, Hardy S, Wendel V, Mehm F, Steinmetz R. Serious games for health: personalized exergames. In: Proceedings of the 18th ACM international conference on Multimedia Pages. New York, NY, USA: ACM Press; 2010 Oct 25 Presented at: International conference on Multimedia Pages; October 25 - 29, 2010; Firenze, Italy p. 1663-1666. [doi: 10.1145/1873951.1874316]
84. Orji R, Mandryk RL, Vassileva J. Improving the Efficacy of Games for Change Using Personalization Models. ACM Trans Comput Hum Interact 2017 Nov 13;24(5):1-22. [doi: 10.1145/3119929]
85. Hagerty BM, Lynch-Sauer J, Patusky KL, Bouwsema M, Collier P. Sense of belonging: a vital mental health concept. Arch Psychiatr Nurs 1992 Jun;6(3):172-177. [doi: 10.1016/0883-9417(92)90028-H]
86. 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]
87. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
88. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
89. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
90. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther 2015 Dec;75:20-31. [doi: 10.1016/j.brat.2015.10.005] [Medline: 26523885]
91. Shechner T, Rimon-Chakir A, Britton JC, Lotan D, Apter A, Bliese PD, et al. Attention bias modification treatment augmenting effects on cognitive behavioral therapy in children with anxiety: randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2014 Jan;53(1):61-71 [FREE Full text] [doi: 10.1016/j.jaac.2013.09.016] [Medline: 24342386]
92. Carlin AS, Hoffman HG, Weghorst S. Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. Behav Res Ther 1997 Feb;35(2):153-158. [Medline: 9046678]
93. Parsons TD, Rizzo AA. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. J Behav Ther Exp Psychiatry 2008 Sep;39(3):250-261. [doi: 10.1016/j.jbtep.2007.07.007] [Medline: 17720136]
94. Klinger E, Bouchard S, Légeron P, Roy S, Lauer F, Chemin I, et al. Virtual reality therapy versus cognitive behavior therapy for social phobia: a preliminary controlled study. Cyberpsychol Behav 2005 Feb;8(1):76-88. [doi: 10.1089/cpb.2005.8.76] [Medline: 15738695]
95. Rizzo A, Parsons TD, Lange B, Kenny P, Buckwalter JG, Rothbaum B, et al. Virtual reality goes to war: a brief review of the future of military behavioral healthcare. J Clin Psychol Med Settings 2011 Jun;18(2):176-187. [doi: 10.1007/s10880-011-9247-2] [Medline: 21553133]
96. Rothbaum BO, Hodges LF, Ready D, Graap K, Alarcon RD. Virtual reality exposure therapy for vietnam veterans with posttraumatic stress disorder. J Clin Psychiatry 2001 Aug 15;62(8):617-622. [doi: 10.4088/JCP.v62n0808]
97. Warner AT, Obrecht A. Standardising humanitarian data for a better response: the Humanitarian eXchange Language. In: HIF/ALNAP Case Study. London: ODI/ALNAP; 2016.
98. Grol R, Grimshaw J. Evidence-based implementation of evidence-based medicine. Jt Comm J Qual Improv 1999 Oct;25(10):503-513. [Medline: 10522231]
99. De Grove F, Breuer J, Hsueh Hua Chen V, Quandt T, Ratan R, Van Looy J. Validating the digital games motivation scale for comparative research between countries. Commun Res Reports 2016 Dec 16;34(1):37-47. [doi: 10.1080/08824096.2016.1250070]

Abbreviations

ABMT: Attention Bias Modification Training
ADHD: attention-deficit/hyperactivity disorder
ANOVA: analysis of variance

ANCOVA: analysis of covariance
BPNS: Basic Psychological Need Satisfaction
CBT: cognitive-behavioral therapy
CCT: computerized cognitive training
IAPS: International Affective Picture System
MTurk: Amazon's Mechanical Turk
RCT: randomized controlled trial
RQ: research question
SDT: Self-Determination Theory
STAI: State-Trait Anxiety Inventory

Edited by G Wadley, M Czerwinski, R Calvo, J Torous; submitted 15.02.18; peer-reviewed by D Coyle, A Andrade; comments to author 27.03.18; revised version received 17.05.18; accepted 06.09.18; published 08.01.19.

Please cite as:
Birk MV, Mandryk RL
Improving the Efficacy of Cognitive Training for Digital Mental Health Interventions Through Avatar Customization: Crowdsourced Quasi-Experimental Study
J Med Internet Res 2019;21(1):e10133
URL: https://www.jmir.org/2019/1/e10133/
doi:10.2196/10133
PMID:30622095

Original Paper

Assessing the Effectiveness of Engaging Patients and Their Families in the Three-Step Fall Prevention Process Across Modalities of an Evidence-Based Fall Prevention Toolkit: An Implementation Science Study

Megan Duckworth[1], BA; Jason Adelman[2], MD, MS; Katherine Belategui[1], RN; Zinnia Feliciano[1], RN; Emily Jackson[2], RN, MBOE, NEA-BC; Srijesa Khasnabish[1], BA; I-Fong Sun Lehman[2], DrPH, MS; Mary Ellen Lindros[3], EdD, RN; Heather Mortimer[1], RN; Kasey Ryan[1], RN; Maureen Scanlan[3], MSN, RN, NEA-BC; Linda Berger Spivack[3], MS, RN; Shao Ping Yu[2], MPH; David Westfall Bates[1,4], MD, MSc; Patricia C Dykes[1,4], RN, PhD, MA

[1]Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, United States

[2]NewYork-Presbyterian Hospital, Columbia University, Manhattan, NY, United States

[3]Montefiore Medical Center, Bronx, NY, United States

[4]Harvard Medical School, Boston, MA, United States

Corresponding Author:
Megan Duckworth, BA
Division of General Internal Medicine
Brigham and Women's Hospital
1620 Tremont Street
3rd Floor Division of General Internal Medicine
Boston, MA, 02115
United States
Phone: 1 6172764527
Fax: 1 6177327072
Email: msd4cs@virginia.edu

Abstract

Background: Patient falls are a major problem in hospitals. The development of a Patient-Centered Fall Prevention Toolkit, Fall TIPS (Tailoring Interventions for Patient Safety), reduced falls by 25% in acute care hospitals by leveraging health information technology to complete the 3-step fall prevention process—(1) conduct fall risk assessments; (2) develop tailored fall prevention plans with the evidence-based interventions; and (3) consistently implement the plan. We learned that Fall TIPS was most effective when patients and family were engaged in all 3 steps of the fall prevention process. Over the past decade, our team developed 3 Fall TIPS modalities—the original electronic health record (EHR) version, a laminated paper version that uses color to provide clinical decision support linking patient-specific risk factors to the interventions, and a bedside display version that automatically populates the bedside monitor with the patients' fall prevention plan based on the clinical documentation in the EHR. However, the relative effectiveness of each Fall TIPS modality for engaging patients and family in the 3-step fall prevention process remains unknown.

Objective: This study aims to examine if the Fall TIPS modality impacts patient engagement in the 3-step fall prevention process and thus Fall TIPS efficacy.

Methods: To assess patient engagement in the 3-step fall prevention process, we conducted random audits with the question, "Does the patient/family member know their fall prevention plan?" In addition, audits were conducted to measure adherence, defined by the presence of the Fall TIPS poster at the bedside. Champions from 3 hospitals reported data from April to June 2017 on 6 neurology and 7 medical units. Peer-to-peer feedback to reiterate the best practice for patient engagement was central to data collection.

Results: Overall, 1209 audits were submitted for the patient engagement measure and 1401 for the presence of the Fall TIPS poster at the bedside. All units reached 80% adherence for both measures. While some units maintained high levels of patient

engagement and adherence with the poster protocol, others showed improvement over time, reaching clinically significant adherence (>80%) by the final month of data collection.

Conclusions: Each Fall TIPS modality effectively facilitates patient engagement in the 3-step fall prevention process, suggesting all 3 can be used to integrate evidence-based fall prevention practices into the clinical workflow. The 3 Fall TIPS modalities may prove an effective strategy for the spread, allowing diverse institutions to choose the modality that fits with the organizational culture and health information technology infrastructure.

(J Med Internet Res 2019;21(1):e10008) doi:10.2196/10008

KEYWORDS

clinical decision support; fall prevention; fall prevention toolkit; health information technology; implementation science; patient safety

Introduction

Falls are a public health problem. Hospitalization increases the risk of falls [1]. In US hospitals, fall rates range from 3.3 to 11.5 falls per 1000 patient days, with about 25% of in-hospital falls resulting in injury [2]. Falls lead to longer lengths of stay, increased costs, and can have severe psychological impacts on patients [3-5].

Fall prevention research has defined the risks that contribute to in-hospital falls and established valid and reliable fall risk assessment tools [6-8]. However, there was a dearth of research regarding fall prevention protocols that link fall risk factors to evidence-based interventions. After identifying this gap, our team interviewed patients who had fallen and their care team members to determine perceptions of why hospitalized patients fall and interventions that could be effective and feasible in the hospital setting [9,10]. The results led to the conclusion that preventing falls in the hospital is a 3-step process as follows: (1) conducting fall risk assessments; (2) developing a tailored fall prevention plan; and (3) implementing that plan consistently along with universal fall precautions. These qualitative results led to the development of the electronic Fall TIPS (Tailoring Interventions for Patient Safety) Toolkit, an intervention that leverages health information technology to provide clinical decision support linking the fall risk assessment to tailored interventions. The Fall TIPS Toolkit was tested in a randomized control trial on >10,000 patients and showed a 25% reduction in fall rates [11]. The results of this study established the linkage between conducting fall risk assessments and implementing tailored, evidence-based interventions to prevent in-hospital falls.

We conducted a case–control study to understand why patients who received the Fall TIPS intervention fell. The most common reason was that patients did not follow their fall prevention plan [12]. Patients often do not believe they are at risk for falls while hospitalized [10]. These data support the hypothesis that simply teaching patients after completing the fall risk assessment and plan is insufficient; patients must be engaged throughout all 3 steps of the fall prevention process. This engagement protocol can improve the partnership with patients for implementing the plan, which is key in further reducing fall with injury rates in hospitals [13].

Despite the widespread adoption of electronic health records (EHRs) in recent years [14], not every health system can adopt the electronic Fall TIPS Toolkit. Barriers include unsophisticated EHR platforms, lack of funds for the toolkit build, and lack of staff engagement to successfully support the roll out. To address the barriers and facilitate spread, our team developed, tested, and iteratively refined a laminated version of the Fall TIPS Toolkit in collaboration with health systems engineers [14,15]. The laminated version of the Fall TIPS Toolkit preserves the clinical decision support of the electronic version by integrating color to provide the linkage between patients' fall risk factors and the evidence-based interventions. It is a low-tech, patient-friendly solution with few barriers to adoption [16]. It is available in both English and Spanish to serve a diverse patient population.

The laminated Fall TIPS Toolkit was evaluated in a 6-month pilot at 2 sites. In a study, adherence to use of the tool was high (>80%), and patient fall with injury rates declined at both sites [13]. This pilot demonstrated the efficacy of the laminated version of the Fall TIPS Toolkit when it is integrated into the workflow and indicates that at least 80% adherence to the Fall TIPS protocol is clinically significant for lowering fall-related injury rates [13].

To further improve the flexibility, adoption, and patient-centeredness of the Fall TIPS Toolkit, we have also developed a patient safety bedside display. This automatically displays each patient's fall prevention plan on the bedside monitor once a nurse has documented the risk assessment and tailored the fall prevention plan in the EHR. This level of automation provides a guarantee that the information displayed at the bedside is up-to-date and is a means of displaying the fall prevention plan in rooms that do not have a visible location to hang the Fall TIPS poster [17].

These 3 modalities, the electronic Fall TIPS Toolkit, the laminated Fall TIPS Toolkit, and the patient safety bedside display, were developed to engage patients in the 3-step fall prevention process and to spread and integrate evidence into practice regardless of an institution's technical capabilities and local factors. The purpose of this study is to assess the effectiveness for engaging patients and family in the 3-step fall prevention process (as defined by patient/family knowledge of their personalized fall risk factors and prevention plan) of each of the Fall TIPS modalities.

Methods

We conducted this study at Brigham and Women's Hospital (Boston, MA, USA), Montefiore Medical Center (MMC; Bronx, NY, USA), and NewYork-Presbyterian Hospital (Manhattan, NY, USA). Each site incorporated the Fall TIPS fall prevention process into practice and built the clinical decision support provided by Fall TIPS into the EHR. At each site, nurses complete the Fall TIPS risk assessment and tailored plan and then documented it in the EHR. Furthermore, the 3 modalities were utilized to present and communicate the patients' fall risk factors and tailored fall prevention plan.

The 3 bedside modalities are as follows: (1) the laminated Fall TIPS poster (Figure 1); (2) electronic Fall TIPS poster (Figure 2); and the paperless patient safety bedside display (Figure 3).

To assess the effectiveness of engaging patients in the 3-step fall prevention process across the Fall TIPS modalities, random audits were conducted asking, "Does the patient/family member know their fall prevention plan?" In addition, random audits measured protocol adherence, defined as the presence of the Fall TIPS fall prevention plan at the bedside.

Weekly data were reported April-June 2017. Fall prevention nurse champions randomly selected, at previously unannounced times, eligible patients or family members for the audit. The inclusion criteria were as follows: patients must be aged ≥18 years; either alert and oriented or have family present and involved in care; English or Spanish speaking; and have a length of stay >24 hours (to allow nurse time to engage patient and family). To conduct the audit, nurse champions verbally asked patients or family members about their knowledge and engagement with their fall prevention plan.

Data were collected on 6 Neurology units and 7 medical or medical-surgical units; these units were chosen for the sample because the team sought to include an analysis of Modality 3 in the study. Modality 3 had only been deployed on 3 Neurology and 2 medical or medical-surgical units at Brigham and Women's Hospital as part of an Agency for Healthcare Research and Quality-funded Patient Safety Learning Lab grant. Neurology and medical or medical-surgical units were then selected for inclusion at MMC and NewYork-Presbyterian Hospital to have equal representation of these services at each hospital. Multimedia Appendix 1 presents a table of the Fall TIPS modality utilized at each hospital by the unit.

The main outcomes measured were the percentage of patients and family members who reported knowing their personal fall risk factors and plan across the 3 Fall TIPS modalities and protocol adherence measured as the display of the personalized fall prevention plan at the bedside.

Figure 1. Modality 1: Laminated Paper Fall TIPS bedside poster, on which a nurse manually documents a patient's fall risks and tailored intervention plan.

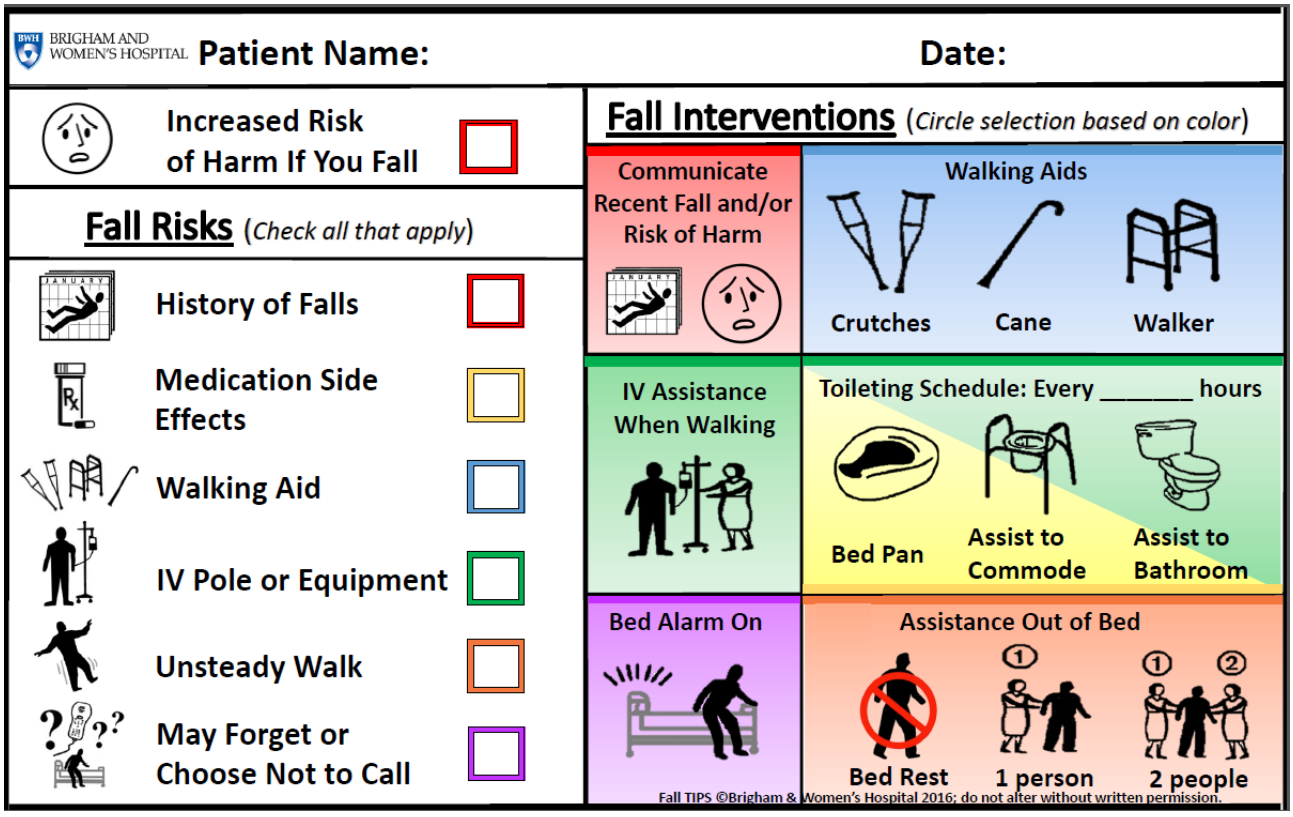


Figure 2. Modality 2: The Electronic Fall TIPS bedside poster, generated by nurse documentation of the personalized fall prevention plan in the electronic health record.

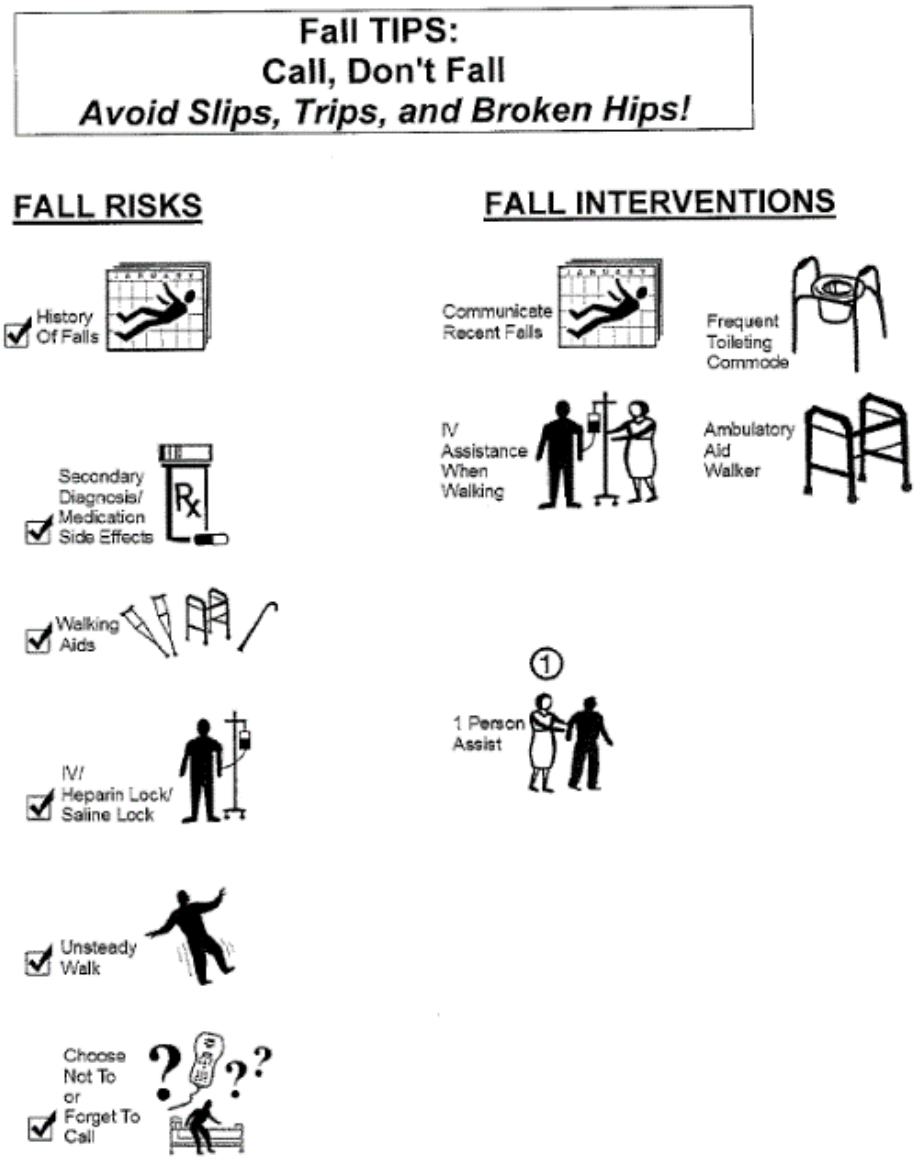
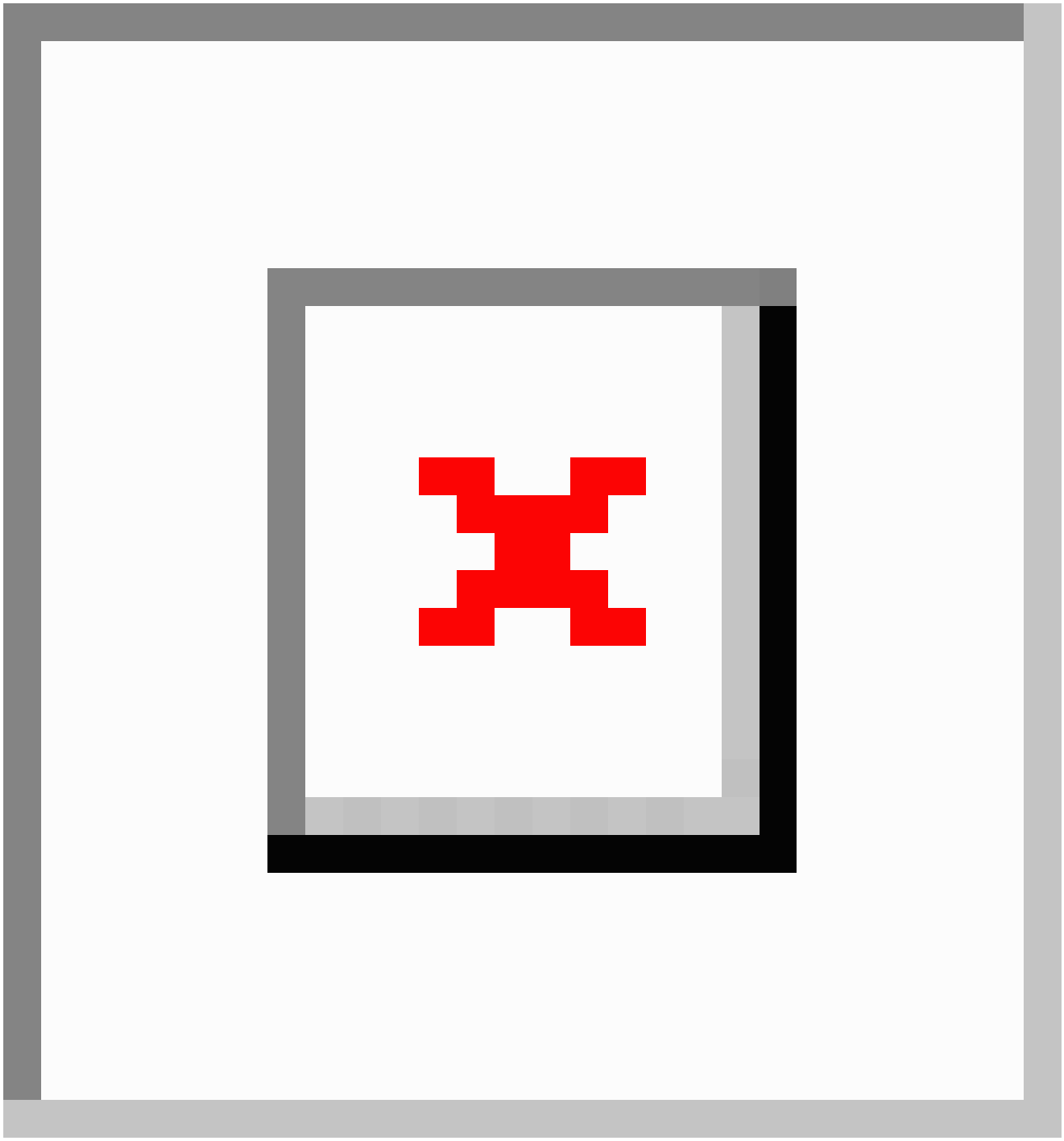


Figure 3. Modality 3: Patient safety bedside display of the Fall TIPS personalized fall prevention plan which is automatically displayed as the screensaver on the computer monitor in the patient's room after nurse documentation in the electronic health record.



Results

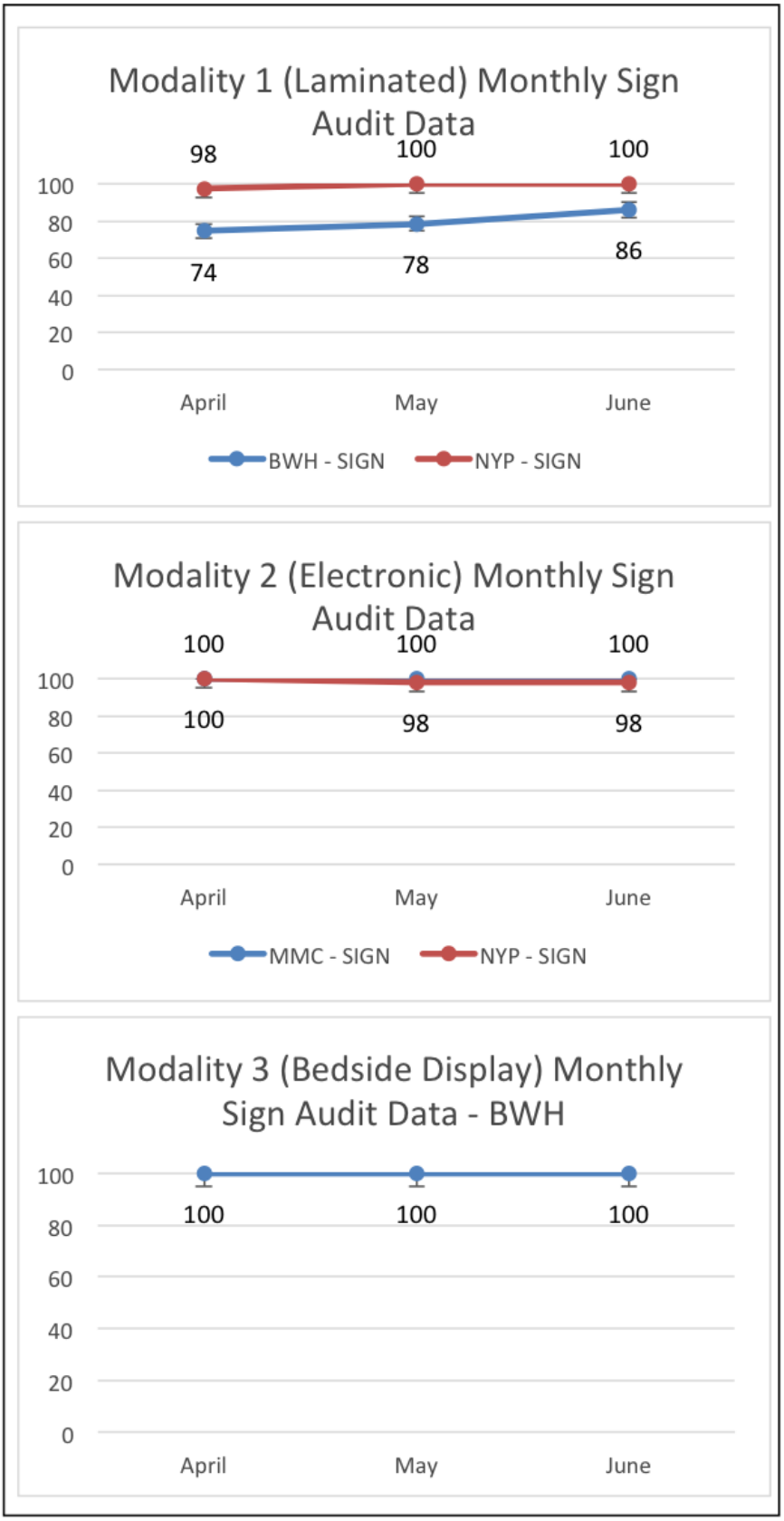
Nurses submitted 1209 audits for the patient engagement measure and 1401 for the presence of the Fall TIPS poster at the bedside. The sample included a diverse population of patients; at MMC, 37.78% (481/1273) of patients reported Hispanic ethnicity, where the Spanish tool is frequently utilized. Table 1 presents the patients' demographics. The average ages of patients at the Brigham and Women's Hospital, MMC, and NewYork-Presbyterian Hospital were 60.5, 60.1, and 60.3 years, respectively. All units reached at least 80% adherence for both measures by the last month of data collection. Figure 4 provides patient engagement audit data over time. Figure 5 provides adherence data to the Fall TIPS protocol as measured by the presence of the personalized fall prevention plan at the bedside. Multimedia Appendices 2 and 3 present audit data counts over time at each site.

Table 1. Patients' characteristics by the site.

Characteristics	Brigham and Women's Hospital (n=1842), n (%)	Montefiore Medical Center (n=1273), n (%)	NewYork-Presbyterian Hospital (n=2582), n (%)
Female	999 (54.23)	709 (55.69)	1343 (52.01)
Race			
American Indian or Native Alaskan	4 (0.22)	3 (0.24)	5 (0.19)
Asian	43 (2.33)	21 (1.65)	76 (2.94)
Black or African American	259 (14.06)	416 (32.68)	310 (12.01)
Declined	13 (0.71)	110 (8.64)	34 (1.32)
Native Hawaiian or Other Pacific Islander	2 (0.11)	1 (0.08)	10 (0.39)
Other	170 (9.23)	509 (39.98)	23 (0.89)
Unavailable	32 (1.74)	25 (1.96)	1433 (55.50)
White or Caucasian	1319 (71.61)	188 (14.77)	691 (26.76)
Ethnicity			
Declined	2 (0.11)	107 (8.41)	33 (1.28)
Hispanic	172 (9.34)	481 (37.78)	361 (13.98)
Non-Hispanic	1593 (86.48)	615 (48.31)	557 (21.57)
Unavailable	75 (4.07)	70 (5.49)	1631 (63.17)

Figure 4. Patient engagement audit data by modality from April to June 2017. BWH: Brigham and Women's Hospital. NYP: NewYork-Presbyterian Hospital; MMC: Montefiore Medical Center.

Figure 5. Adherence to the Fall TIPS protocol, as measured by the presence of the personalized fall prevention plan at the bedside, by modality from April to June 2017. BWH: Brigham and Women's Hospital; NYP: NewYork-Presbyterian Hospital; MMC: Montefiore Medical Center.



Discussion

Principal Findings

This study sought to determine if there was variability in the ability to engage patients in the 3-step fall prevention process across the Fall TIPS modalities. The results illustrate little difference in the ability to engage patients across the 3 modalities. All units, regardless of the modality and site, reached clinically significant rates (>80%) of patient engagement and adherence with the sign protocol to reduce fall and fall with injury rates [13]. This suggests that each Fall TIPS modality is effective at engaging patients in the 3-step fall prevention process and so can be used to implement the evidence into practice.

The different levels of automation provide flexibility for institutions to individualize their Fall TIPS implementation approach. Institutions must assess which modality is most appropriate, considering factors like EHR capability, the commitment of health information technology support, financial constraints and local geographic realities, such as room layouts, or whether there are uniform bedside monitors. Further research investigating how each Fall TIPS modality impacts fall and fall with injury rates is needed. In practice, Fall TIPS is used in >100 hospitals and continues to spread. Interested hospitals have free access to the Fall TIPS Fall Prevention Toolkit and training materials through the Fall TIPS Collaborative.

Limitations

As Fall TIPS was implemented at 3 different institutions, differences in the communication channels, social systems, the support from leadership and the timing of Fall TIPS implementation pose limitations to this study. Some of these factors are discussed elsewhere [13]. Each of these factors could have confounded the levels of patient engagement with fall prevention as related to the Fall TIPS modality utilized and protocol adherence.

Finally, this is an implementation science study. The study was not designed to randomize units to each modality but to assess the efficacy of the Fall TIPS modalities within the existing institutional frameworks. This was a study of the uptake of evidence-based practice across modalities. The implementation of practice into the workflow does not allow for perfect comparability but demonstrates that achieving clinically significant rates of adherence to an evidence-based fall prevention program in the workflow is possible.

Conclusions

Fall TIPS is an evidence-based fall prevention intervention that provides built-in clinical decision support to engage patients and family in the 3-step fall prevention process. It has been iteratively developed and refined to include 3 modalities with varying degrees of automation while preserving the clinical decision support inherent to Fall TIPS. The 3 modalities provide flexibility for health systems with different capabilities to integrate the fall prevention evidence into practice.

This study demonstrates that across the 3 modalities, the laminated Fall TIPS Toolkit, the electronic Fall TIPS Toolkit, and the patient safety bedside display of Fall TIPS, there is no significant clinical difference in ability to engage patients. Previous research has shown that patient engagement is the crux of improving fall rates and that at least 80% adherence with the Fall TIPS protocol is clinically significant for doing so [13]. Therefore, the ability to engage patients corresponds with the efficacy of the Fall TIPS modalities. Overall, this study suggests that providing 3 Fall TIPS modalities is an effective and flexible approach for promoting adoption and spread.

Acknowledgments

This study is funded by the Agency for Healthcare Research and Quality (Grant #1-P30-HS023535).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Fall TIPS modality utilized at each hospital by unit.

[PNG File, 21KB - jmir_v21i1e10008_app1.png]

Multimedia Appendix 2

Monthly engagement audit data by modality at each site.

[PNG File, 22KB - jmir_v21i1e10008_app2.png]

Multimedia Appendix 3

Monthly protocol adherence audit data by modality at each site.

[PNG File, 21KB - jmir_v21i1e10008_app3.png]

References

1. Lakatos BE, Capasso V, Mitchell MT, Kilroy SM, Lussier-Cushing M, Sumner L, et al. Falls in the general hospital: association with delirium, advanced age, and specific surgical procedures. Psychosomatics 2009;50(3):218-226. [doi: 10.1176/appi.psy.50.3.218] [Medline: 19567760]
2. Bouldin ELD, Andresen EM, Dunton NE, Simon M, Waters TM, Liu M, et al. Falls among adult patients hospitalized in the United States: prevalence and trends. J Patient Saf 2013 Mar;9(1):13-17 [FREE Full text] [doi: 10.1097/PTS.0b013e3182699b64] [Medline: 23143749]
3. Haines TP, Hill AM, Hill KD, Brauer SG, Hoffmann T, Etherton-Beer C, et al. Cost effectiveness of patient education for the prevention of falls in hospital: economic evaluation from a randomized controlled trial. BMC Med 2013 May 22;11:135 [FREE Full text] [doi: 10.1186/1741-7015-11-135] [Medline: 23692953]
4. Krauss MJ, Nguyen SL, Dunagan WC, Birge S, Costantinou E, Johnson S, et al. Circumstances of patient falls and injuries in 9 hospitals in a midwestern healthcare system. Infect Control Hosp Epidemiol 2007 May;28(5):544-550. [doi: 10.1086/513725] [Medline: 17464913]
5. Jefferis BJ, Iliffe S, Kendrick D, Kerse N, Trost S, Lennon LT, et al. How are falls and fear of falling associated with objectively measured physical activity in a cohort of community-dwelling older men? BMC Geriatr 2014 Oct 27;14:114 [FREE Full text] [doi: 10.1186/1471-2318-14-114] [Medline: 25348492]
6. Tideiksaar R, Feiner CF, Maby J. Falls prevention: the efficacy of a bed alarm system in an acute-care setting. Mt Sinai J Med 1993 Nov;60(6):522-527. [Medline: 8121429]
7. Schwendimann R, Milisen K, Bühler H, De Geest S. Fall prevention in a Swiss acute care hospital setting Reducing multiple falls. J Gerontol Nurs 2006 Mar;32(3):13-22. [Medline: 16544453]
8. Cumming RG, Sherrington C, Lord SR, Simpson JM, Vogler C, Cameron ID, Prevention of Older People's Injury Falls Prevention in Hospitals Research Group. Cluster randomised trial of a targeted multifactorial intervention to prevent falls among older people in hospital. BMJ 2008 Apr 05;336(7647):758-760 [FREE Full text] [doi: 10.1136/bmj.39499.546030.BE] [Medline: 18332052]
9. Dykes PC, Carroll DL, Hurley AC, Benoit A, Middleton B. Why do patients in acute care hospitals fall? Can falls be prevented? J Nurs Adm 2009 Jun;39(6):299-304 [FREE Full text] [doi: 10.1097/NNA.0b013e3181a7788a] [Medline: 19509605]
10. Carroll DL, Dykes PC, Hurley AC. Patients' perspectives of falling while in an acute care hospital and suggestions for prevention. Appl Nurs Res 2010 Nov;23(4):238-241 [FREE Full text] [doi: 10.1016/j.apnr.2008.10.003] [Medline: 21035035]
11. Dykes PC, Carroll DL, Hurley A, Lipsitz S, Benoit A, Chang F, et al. Fall prevention in acute care hospitals: a randomized trial. JAMA 2010 Nov 03;304(17):1912-1918 [FREE Full text] [doi: 10.1001/jama.2010.1567] [Medline: 21045097]
12. Dykes PC, I-Ching EH, Soukup JR, Chang F, Lipsitz S. A case control study to improve accuracy of an electronic fall prevention toolkit. AMIA Annu Symp Proc 2012;2012:170-179 [FREE Full text] [Medline: 23304286]
13. Dykes PC, Duckworth M, Cunningham S, Dubois S, Driscoll M, Feliciano Z, et al. Pilot Testing Fall TIPS (Tailoring Interventions for Patient Safety): a Patient-Centered Fall Prevention Toolkit. Jt Comm J Qual Patient Saf 2017 Aug;43(8):403-413. [doi: 10.1016/j.jcjq.2017.05.002] [Medline: 28738986]
14. Katsulis Z, Ergai A, Leung WY, Schenkel L, Rai A, Adelman J, et al. Iterative user centered design for development of a patient-centered fall prevention toolkit. Appl Ergon 2016 Sep;56:117-126. [doi: 10.1016/j.apergo.2016.03.011] [Medline: 27184319]
15. Zuyev L, Benoit AN, Chang FY, Dykes PC. Tailored prevention of inpatient falls: development and usability testing of the fall TIPS toolkit. Comput Inform Nurs 2011 Feb;29(2):93-100 [FREE Full text] [doi: 10.1097/NCN.0b013e3181f9dbe9] [Medline: 20975543]
16. Leung WY, Adelman J, Bates DW, Businger A, Dykes JS, Ergai A, et al. Validating Fall Prevention Icons to Support Patient-Centered Education. J Patient Saf 2017 Feb 22:-. [doi: 10.1097/PTS.0000000000000354] [Medline: 28230576]
17. Duckworth M, Leung E, Fuller T, Espares J, Couture B, Chang F, et al. Nurse, Patient, and Care Partner Perceptions of a Personalized Safety Plan Screensaver. J Gerontol Nurs 2017 Apr 01;43(4):15-22. [doi: 10.3928/00989134-20170313-05] [Medline: 28358972]

Abbreviations

EHR: electronic health record
MMC: Montefiore Medical Center
TIPS: Tailoring Interventions for Patient Safety

Edited by G Eysenbach; submitted 05.02.18; peer-reviewed by M Immonen, H Hawley-Hague, D Lipardo; comments to author 15.08.18; revised version received 13.09.18; accepted 03.10.18; published 21.01.19.

Please cite as:
Duckworth M, Adelman J, Belategui K, Feliciano Z, Jackson E, Khasnabish S, Lehman IFS, Lindros ME, Mortimer H, Ryan K, Scanlan M, Berger Spivack L, Yu SP, Bates DW, Dykes PC
Assessing the Effectiveness of Engaging Patients and Their Families in the Three-Step Fall Prevention Process Across Modalities of an Evidence-Based Fall Prevention Toolkit: An Implementation Science Study
J Med Internet Res 2019;21(1):e10008
URL: http://www.jmir.org/2019/1/e10008/
doi:10.2196/10008
PMID:30664454

Original Paper

Farm Owners and Workers as Key Informants in User-Centered Occupational Health Prototype Development: A Stakeholder-Engaged Project

Bryan Weichelt[1], PhD, MBA; Casper Bendixsen[1], PhD; Matthew Keifer[2], MD MPH

[1]Marshfield Clinic Research Institute, National Farm Medicine Center, Marshfield, WI, United States

[2]University of Washington and Puget Sound Veterans Administration, Seattle, WA, United States

Corresponding Author:
Bryan Weichelt, PhD, MBA
Marshfield Clinic Research Institute
National Farm Medicine Center
701 W Kalsched Ave
Marshfield, WI, 54449
United States
Phone: 1 715 221 7276
Fax: 1 715 221 3535
Email: weichelt.bryan@mcrf.mfldclin.edu

Abstract

Background: The cost of workplace injuries and illnesses significantly impacts the overall cost of health care and is a significant annual economic burden in the United States. Many dairy and pork farm owners in the Upper Midwest have expanded operations and taken on the role of manager and employer yet receive little training in injury prevention, farm safety, or workers' compensation programs and processes. Clinicians play a key role in the return to work of injured and ill farmers and farmworkers to their jobs, though little to no formal training is offered by medical schools.

Objective: This stakeholder-engaged project aimed to develop a prototype application designed to assist clinicians in returning injured farmworkers to light-duty job assignments with their current employers and to assess farm owners' and managers' attitudes toward and barriers to adopting mobile health tools for themselves or their employees.

Methods: We conducted 12 semistructured interviews with English-speaking farm owners and farmworkers from the Upper Midwest: 5 English-speaking and Spanish-speaking farmworker focus groups and 8 postproject interviews with farm owners that focused on attitudes and barriers to adoption of the developed software. Interviews and focus groups were audio recorded, and data were analyzed and thematically coded using audio coding.

Results: Interviews and worker focus groups guided an iterative design and development cycle, which informed workflow design, button placement, and output sheets that offer specific light-duty farm work recommendations for the injured worker to discuss with his or her employer.

Conclusions: The development of a complex prototype intended to impact patient care is a significant undertaking. Reinventing a paper-based process that can eventually integrate with an electronic health record or a private company's human resources system requires substantial stakeholder input from each facet including patients, employers, and clinical care teams. The prototype is available for testing, but further research is needed in the form of clinical trials to assess the effectiveness of the process and the software's impact on patients and employers.

(J Med Internet Res 2019;21(1):e9711) doi:10.2196/jmir.9711

KEYWORDS

agriculture; farmworkers; injuries; occupational medicine; return to work; software application

Introduction

The cost of workplace injury and illness represents a significant factor in the overall cost of health care and the cost of doing business in the United States. Work-related injuries account for approximately 30% of the total injury burden in the United States among working-age people [1]. Nationally, Leigh and colleagues estimated that total direct and indirect costs for work-related injury and illness were US $1,555.5 billion [2]. Workers and their families bore the highest percentage of that cost. Indirect costs at an estimated US $103.7 billion result from a combination of loss of wages and fringe benefit costs, loss of home productivity, and slowed or stopped production due to a key missing employee or replacement [2].

Accurate, up-to-date workplace injury estimates would be difficult to develop from administrative data alone, as the majority of agricultural operations do not contribute information to the 2 common sources of injury and illness data: Occupational Safety and Health Administration 300 reporting [3] and worker compensation systems, which vary greatly across US states. Most agricultural operations have fewer than 11 employees and, thus, are not required to file Occupational Safety and Health Administration 300 forms, which are the source of the Bureau of Labor Statistics data. Also, in many states, worker compensation rules exempt all but the largest agricultural employers from purchasing worker compensation [4]. Based on studies previously done in the Midwest [5,6] and research from work done on dairy and pork producers in other geographic areas [7-10], it is clear that these 2 industries endure significant injury and illness among their respective workforces.

The growth of concentrated animal feeding operations has been a reality in modern agriculture for the past 2 decades [11]. As family farms expand or fold, more agricultural operations are hiring workers who face the risks inherent in the agricultural workplace [12]. In Wisconsin and elsewhere, this rural workforce has rapidly diversified to include more immigrant and Hispanic workers, primarily from Mexico [13,14]. It is estimated that approximately 62% of milk produced in the United States is produced using immigrant labor [15]. Furthermore, the undocumented status of immigrant workers is, in itself, an occupational hazard [16]. This is further exacerbated by the already hazardous work environment, which remains among the most dangerous US industries. For example, 2015 data from the National Institute for Occupational Safety and Health shows that young workers (age<18 years) in agriculture were 44 times more likely to die on the job when compared to all other industries combined [17].

As health care expands and health insurance becomes available to more people through the Affordable Care Act [18], agricultural injuries to workers as well as farm owners will increasingly be cared for and managed by primary care practitioners (nurse practitioners, primary care physicians, and physician's assistants) who must manage returning these workers to the workplace. However, despite the frequency with which clinicians are faced with managing the return to work (RTW) process, most have little training in the skill [19]. There also exist barriers even for those who are competent in this skill. Guzman et al [20] found that Canadian physicians perceived the willingness of the workplace to accommodate the injured worker as being the second most important factor in facilitating RTW after patient pain perception.

Programs that facilitate the early and safe return of recovering workers to some level of function in the workplace may substantially reduce employer costs and also be of significant benefit to the worker [21]. There are many factors that influence the success of early RTW efforts [22]. Among these are the perceived self-worth of the worker [23], the worker's preinjury job satisfaction, pain severity, worker age, clinician expectations of RTW, employer attitudes, and competency of the RTW coordinator, who is typically a clinical staff member assisting with the process and serving as the linchpin between the patient's physician and the employer [24,25]. Some of the most important factors relate to assuring that the accommodation of returning workers is efficient and effective, and legal and privacy constraints are recognized and mitigated [21,26].

In certain professions, the variability of work tasks is fewer than others. Within agriculture, despite the differences in processes and technology, the raising of large animals (pigs and cows) includes several necessary activities that are indispensable (feeding, manure management, milking, transport, and animal health management). The physical variability of tasks among these activities is relatively limited. Thus, it was feasible to develop a robust and versatile compendium of ergonomically characterized tasks in large animal agriculture to guide clinicians in returning injured workers through transitional work to a full recovery.

The RTW process may involve several key players. A clinician will generally manage the worker medically. A physical therapist, occupational therapist, or specially trained RTW professional may oversee the actual RTW (uncommon with small employers and rural practices), and the employer will interpret RTW restrictions from the clinician. RTW limitations are conveyed to the employer in a generic, simple list of the patient's physical limitations. This approach is correct in concept but flawed in design, at least in the RTW of agricultural workers. In concept, work involves a limited number of functions: lifting, bending, standing, pulling, pushing, etc. While clinicians may be able to estimate a worker's limitations and describe those on an RTW guidance sheet, they generally know little about what the workplace tasks truly involve, and most employers are not trained to interpret these abstract documents in the context of the workplace.

At the onset of this project, no known products existed privately or commercially that specifically addressed the RTW facilitation for agricultural employment. This project was designed to replace this speculative, abstract document with a concrete directive through a task-based software application fed from a database of occupational work task ergonomics from dairy and pork work. The system was expected to both facilitate an employer's ability to adapt it to their own worksite and educate the clinician about the worksite. Additional benefits may include enhanced worker understanding and participation in the RTW decision making and improved communication between employer and clinician.

Methods

Project Aims

The primary goal of this 5-year project was to develop a compendium of tasks in the dairy and pork industries that encompasses the various types of production work and to redistribute task information through a software application to assist clinicians and employers in RTW processes for injured workers. First, we conducted key informant interviews with experts in dairy and pork production to identify the main work processes and specific work tasks that make up these processes. Second, we conducted a series of farmworker focus groups in parallel with farm owner or manager interviews, primarily to usability test the initial mock-up designs of the new RTW system interface. Lastly, we conducted follow-up interviews with farm owners or managers regarding their use of technologies, including mobile health-related tools for themselves or to manage the health and safety of their employees.

Through the first 2 years of this project, information on the tasks that make up work in dairy and pork production was collected. In a series of in-office meetings, we shared initial findings with research staff about the different processes that make up dairy and pork production and the different injury issues faced by the workforce. This information was obtained through the guidance and assistance of project advisors from the Professional Dairy Producers of Wisconsin, the largest dairy organization in Wisconsin; the Pork Board; and others who were interviewed during the coordinated key informant process (year 1) described below. This process also educated our team occupational therapists, who were tasked with collecting ergonomic data (years 1-2), about the terminology necessary for effective communication with dairy and pork producers.

Coordinated Key Informant Interview Process

There were 3 separately funded occupational health research projects running in parallel that all employed key informant interviews at the start of their project activities. The 5-year projects coordinated their interviews to allow an economy of scale in these data collection processes, saving funds and minimizing demands on busy advisees and respondents. Each of the 3 projects identified key informants individually. Project personnel identified gaps in the list and suggested additional informants to be interviewed. There were 6 farm owners with employees who participated in these initial interviews. Of these participants, 3 were male and 3 were female. Questions were developed, pilot tested, and reviewed by each of the 3 project teams based on their projects' information needs. This allowed the assembly of questions tailored to the key informant being interviewed and inclusive of the questions that each of the projects wanted to be addressed, while not overburdening participants with multiple interviews.

Ergonomic Data Collection

The project team employed a modified version of DSI Work Solutions, Inc.'s unpublished methodology for collecting ergonomic measurements of the functional job assessments of work tasks. The codeveloper of the decision support initiative (DSI) methodology (and consultant to the project) worked with the authors to develop a nonproprietary tool that can remain for use in the public domain to collect information on functional job assessments (or analysis) in pork and dairy production. This tool was designed to allow ease in transferring data to the RTW guidance software application database. Functional job assessments or job demand analyses exist for many jobs in nonagricultural settings. This project took steps to develop these profiles for common tasks in dairy and pork which are now available within the prototype app. Lessons learned from the collection and integration of ergonomic data into a software application will be discussed in another manuscript.

Focus Groups

Farmworkers were recruited for focus groups from among Anglo and Hispanic workers through our participant dairy and pork workplaces in Wisconsin. These farmworkers were asked to review and interpret RTW instructions. English worker instructions were translated into Spanish by a bilingual research specialist and reviewed by Marshfield Clinic Health System (MCHS) interpreters.

Instructions for RTW specialists were tested with occupational medicine providers at MCHS. Clinicians at MCHS in primary care and occupational medicine were recruited to review and offer advice on the forms designed for guiding clinicians in returning a worker to light-duty job activity. The results from the clinician interviews will be described in another paper.

The objectives of the focus groups and interviews were 3-fold: (1) to assess participants' perception of the current process of returning an injured farmworker back to work, potentially in a limited capacity, and to see where improvements could be made; (2) to gauge the acceptance of the concept of suggesting potential tasks along with the physical limitations and treatment instructions normally included on the worker's compensation report; and (3) to capture these data from farm employers and their employees, specifically including data from the Spanish-speaking workers—a vulnerable, yet growing population of workers in the Upper Midwest dairy industry. Feedback from the focus groups and interviews were used to help steer the initial designs and the iterative development of the software and its outputs. Focus groups have shown to open up discussion and conversation, more so than just one-on-one interviews, particularly in software design but also among Spanish-speaking Wisconsin dairy workers [27,28].

We conducted 5 farmworker focus groups on Wisconsin dairy farms; 3 were Spanish-speaking, and 2 were English-speaking groups. The focus groups took place over the participants' lunch hour or at a time their schedule allowed. Participants were provided with a box lunch and also received a US $20 gift card. A usability analyst facilitated the focus groups in a private area without the employer or manager present. An interpreter was present for the Spanish-speaking groups to translate all questions and answers for the benefit of the usability analyst and the participants. The usability analyst presented the existing paper-based workers' compensation form used in clinical practice at the MCHS and facilitated a semistructured discussion regarding participants' experience with existing RTW processes. The analyst also introduced the concept of a software application

that would utilize an algorithm to present potential work tasks based on the physical limitations presented by the injury, asking questions regarding the design, button placement, and terminology. Paper copies of early design mockups of the application interface were handed out and referred to during the process (Figure 1).

Participants in each of the 5 focus groups were asked a series of questions from a semistructured questionnaire; these questions focused on elements of the Workers' Compensation process including past injuries, experiences, and feedback on the detailed variables of the new forms' mock-up designs.

Interviews

We conducted 6 interviews with Wisconsin dairy farm owners or managers to assess the current practices and level of knowledge and to gather insight on initial conceptual designs of a proposed output sheet from the SafeReturnToWork.org (SRTW) app. In total, 2 participants who were owners and 4 were farm managers or identified as administrators; 3 were male and 3 female. All 6 farms employed nonfamily full-time workers. The workers' compensation form was a standard form that MCHS was using at the time of this study. The initial conceptual design was referred to and guided the semistructured farmer interviews (see Figure 2).

Analysis

Interviews and focus groups were audio recorded. The primary method of analysis was audio coding [29]. The analyst and authors used inductive analyses to identify themes and organize them into thematic sections to inform the iterative software development cycle, as highlighted by participants' direct quotes in the results section below.

Figure 1. Early mock-up of the application interface.

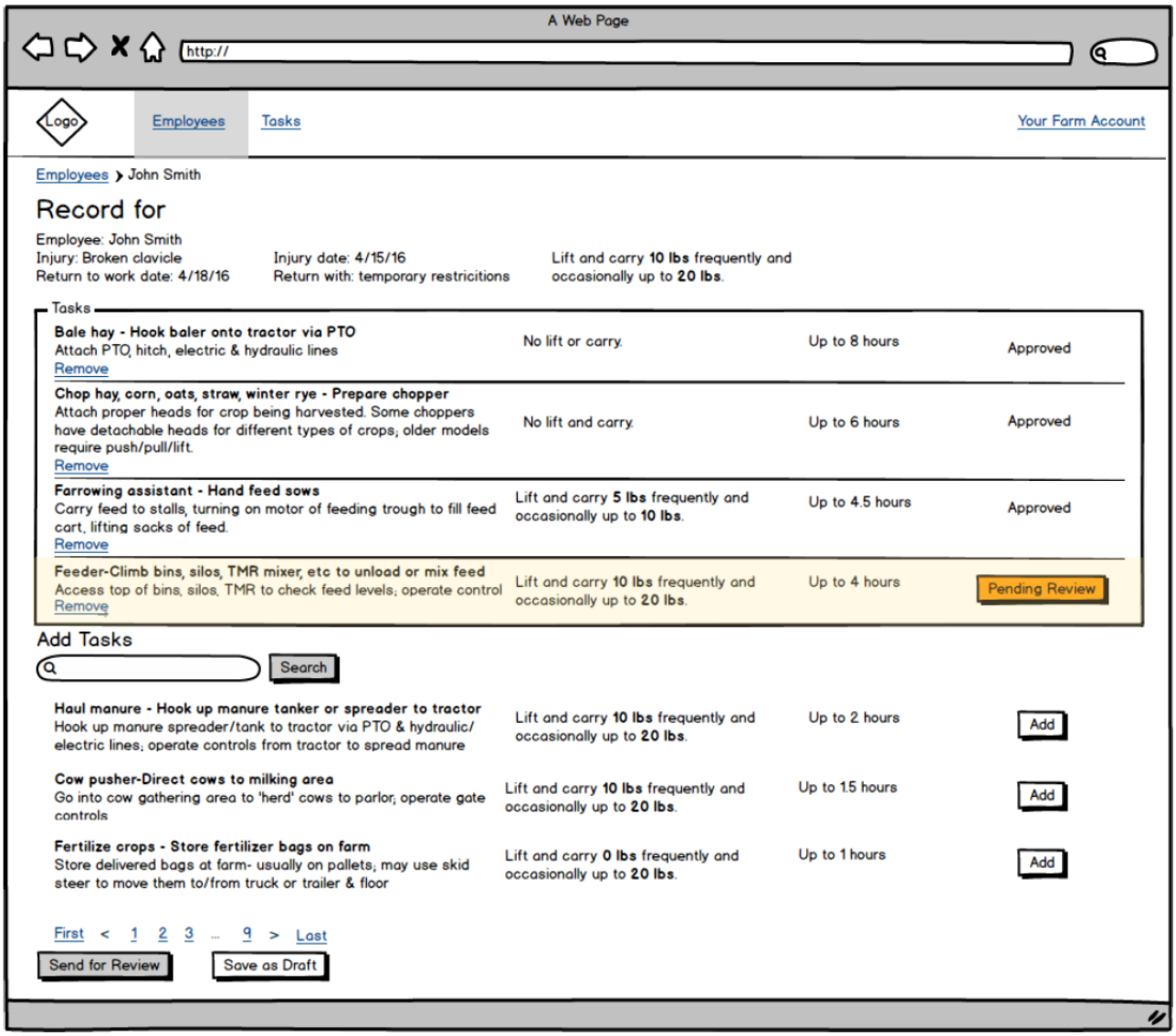


Figure 2. Photo of an interview with a farm owner or manager using the SafeReturnToWork.org application.

Results

Participants' Experience with the Workers' Compensation Process and Claims

A substantial majority of the participants had some knowledge and experience regarding the workers' compensation process. There was 1 farmer who was knowledgeable about the costs associated with an injury claim and spoke specifically about the multiplier effect of accepting wages. The participant stated the following:

> *the wages can kill you. You can have an awful lot of medical expenses, but compared to wages...We would have been way better off on the experience mod, if we did not take any wages*
>
> *You know with all [name omitted]'s hospitalization, and all the surgeries on his foot and the physical therapy, uhm, and the minimal amount of wages that we took, you know, someone estimated to me that the wages probably were ten to twenty-fold impact. It was unbelievable. When they do your whole mod, this long long calculation, wages apparently are weighted, in those calculations, significantly more than medical. And I just had another person tell me too, as he was showing me their lost claims on workers' comp, but they had a way lower mod than us, and he said' because he's really trying not to have lost wages. He said if you can find something for them to do, you're way better off than taking the lost wages.*

Other participants discussed their experience in managing worker claims by discussing complexities, challenges, and the accommodations made. There was 1 farmer who noted specific challenges, namely variation in available work options due to seasonality. The dairy industry is a year-round operation. Thus, many workers are immigrants, not migrant workers, meaning they have immigrated (moved permanently) to the area rather than visiting for temporary employment. This is a result of the demand for steady, year-round labor. Cows on dairy operations are milked 2 to 3 times per day, every day of the year. Some operations only shut down long enough to clean and wash the area, then start milking another section of the herd. Even with the year-round labor demand, there is still some fluctuation in the farm tasks. Spring (tillage and planting), summer (hay harvest), and fall (grains harvest) are still the busiest seasons in terms of the hours of labor needed to operate a successful dairy. Thus, winter has a lull in available light-duty job assemblies for injured workers. A participant mentioned the following:

> *In the summer we're a little more flexible, but in the winter we don't have that flexibility. So that was our one big one, and it was a winter injury, we just didn't have the work that was matching, so. The summer is*

much easier for us, I'm sure it depends on each business.

Another farmer less familiar with the process noted that their operation just does not have as much flexibility in positions. He stated the following:

The conundrum is that we don't really have any work that's sedentary. The very light work is very minimal. You know, carrying ten pounds, not bending over, that kind of thing, we just don't have lot of that work. Even light work, you know it's, they're lifting their hands above their waste, they're reaching, all those kinds of things put them in those categories where they can't really do a whole lot for us

When discussing the topics of seeking care and the number of claims filed, a discussion with 1 farmer led into the mix of culturally-based decision making in regard to the appropriateness of seeking care for subjectively minor injuries. During an interview with a farmer who employs English-speaking and Spanish-speaking workers on the dairy operation, 1 incident, in particular, was highlighted. The participant noted that all the claims the farm has had have come from English-speaking workers and went on to say the following:

The Hispanics, uh, they uh, they're pretty careful about that stuff. I mean ah going to the doctor if they're injured on the job. They're real fix-it-myself. We had one guy that cut his leg, and uhm didn't want anyone to know, one of the employees told me, so I went over to talk with him. and I said you cut your leg, he said yup, I know how to take care of my leg, I know that when I get home tonight I'm going to elevate it. I don't need stitches. I said I need to see it and he showed it to me and it didn't need stitches. He said I know how to clean a wound, I know how to take care of it, I don't need to go to a doctor. Ok. You know, it wasn't one that needed stitches. If it needed stitches it would be a little different story. But they're pretty careful about that stuff.

Initial Conceptual Designs of the SafeReturnToWork Software Interface

There was a mix of responses among farm owner or manager interview participants when reviewing the conceptual designs. This was also the most involved and time-consuming section of the interview, where participants delved into specific sections of the output sheet (eg, physician contact information, checkboxes vs narratives, suggested tasks, limitations, medications, and treatment plan).

When reviewing the section regarding physician contact information, 1 participant was very quick to respond. Its usefulness was questioned immediately, based on the experience of this farmer and the perception of the transpired events. The interviewee stated the following:

I do not think a doctor will talk to us. I don't think it matters. HIPAA. They won't talk, I can't call and ask about my children that are 21 years old, why would they allow me to call and talk to them about an injured farm worker?

We haven't had, I can think of two of them, where we tried, and we didn't get. Maybe it would be just the contentious nature of their injury and those two individuals were you know, I don't know. but I did not, those two we did try and we just decided that we cannot do that. So we just went back to our insurance company and had them do what they can on that end to try and get somewhere.

Interviewer: So when you tried to call the doc, did they refuse to speak with you?

yea, they wouldn't call back"

Finally, farm owners or managers were asked if they thought Web-based resources would be useful for an injured farmworker and if they would access those resources. When asked about thoughts on using the internet to access therapy- and recovery-related resources to aid in the rehabilitation process of an injured patient, one farmer responded by stating the following:

For me, that would not be customer service at all. Give them the information in hard copy, I think is the customer service route. Because I don't think if I'm partially laid up, and I've been injured at work, that I should have to come home, search the internet, find the website, find the information, print it out. I'm not going to do the work. If it's sitting around on the counter, my kids or my spouse will probably say, hey your doctor gave you this, you better get going on this. If I have to go to a computer and find it, I don't think it's ever gonna get looked at, we're done

During these farm owner or manager interviews, further input was offered on a number of related topics, including some foresight into the usefulness of a system to inform future health and safety interventions. There was 1 such farmer who was specifically interested in the outcomes of the Safe RTW project in tracking data and advising the industry on best practices that will lead to fewer injuries, fewer claims, and cost savings.

I see the end result to say, workers comp rates on dairy farms – here are the injures – here are the bulk of injuries that farms have with workers comp. What can we do as an industry to minimize these high number injuries, from a client perspective, to bring down our work comp rates? That's the other side of it, as an industry, what can we gain by those types of things, by knowing those types of things. Because in five years if you guys go back to your data warehouse and say the number one workers comp [claim] is eye injuries or wrist injuries…how do these accidents happen and what causes those injuries. You guys as dairy farmers, if you eliminate these situations how much less work comp losses would you have on your bottom line?

Focus Groups with Farmworkers

A total of 35 farmworkers participated in the 5 focus groups comprising 3 Spanish-speaking and 2 English-speaking groups (see Table 1). A total of 32 males and 3 females participated. Of the farmworkers, 20 identified as Hispanic/Latino, 14

identified as white or Caucasian, and 1 did not self-identify. Spanish was primarily spoken by 20 workers. The project team did not collect any further demographic data from these participants. However, it is likely that the Spanish-speaking participants had similar home states, limited prior agricultural work experience, and similar current jobs in dairy (milkers, pushers, and feeders) as reported by Liebman et al [16].

These focus groups produced significant design changes to the prototype light-duty job activity forms, which shifted to a design that closely resembles the current workers' compensation form used throughout the MCHS.

SafeReturnToWork Adoption

On-farm interviews were conducted with farm employers, inquiring about the possible use of this technology and the concept of engaging with physicians about workers' compensation related issues. In the final year of this project, 8 farmers were interviewed to further explore barriers to adoption of the SRTW system and its mobile format in regard to the use of mobile health-related tools. Interviewees included married couples, single farmers, and farm managers. All participants had previously engaged with research staff regarding the SRTW system at different points in its development. The interviews were specifically focused on how farmers have communicated with physicians regarding injured employees in the past and how the SRTW system could aid and improve that communication. Discussions gravitated toward the strengths and weaknesses of the paperwork an injured worker would return with from the physician, the attractiveness of having the same information over the Web in a more useful format, and how farmers and farm managers would prefer to communicate during a light-duty work period. Interview and data analyses were conducted using the same methods noted in this paper.

Sharing a specific story, 1 farmer articulated a common theme regarding the difficulties of communicating what a light-duty regimen could consist of with a physician and why it is necessary to the worker. This particular worker was injured by a dairy cow and sought medical care for a broken rib and punctured lung. The worker was assigned a 45-day light-duty requirement from the physician. However, in as early as 20 days, the worker was asking to come back to work, afraid of either being fired or fearing there would be repercussions regarding his work in the future, despite the employer's reassurance. Similar worker fears were uncovered through focus groups by Liebman and colleagues [16]. The farmer felt the physician could have aided him more in explaining why a light-duty regimen was necessary. The farmer reported 2 written communications from the physician, both of which he felt were difficult to understand and were uninterpretable by the worker, so he opted for a more restrictive regimen. For example, while work in a skid steer would have suited the employee's skills, experience, and the light-duty work orders, the farmer felt that the climbing in and out of the skid steer was beyond the recommendations. He agreed with the interviewers that he would have utilized the proposed version of the SRTW system to find more jobs for the worker and to further explain the possible tasks to the physician.

Another farm couple with a similar story stated the following:

> *We're probably making assumptions about [the physician's] job and we know he's making assumptions about ours. […] It couldn't hurt to be able to provide education both ways.*

All interviewees agreed that the proposed enhancements to the system would be helpful to farms with many employees (most participants had 15 or more). They also felt that workers' compensation insurance carriers should be considered users and financiers of such a technology, making the argument that it was their business to keep a client's employees healthy during a light-duty assignment.

Unexpected Outcomes and Deliverables

The interview findings led the research team into an unplanned investigation of workers' compensation calculator variation across the Upper Midwest. The results of those investigations are discussed in another manuscript. The team also created an informational page on the SRTW website specifically designed to educate farm owners and managers of the costs related to workers' compensation claims for agricultural operations. That information was reviewed by 3 representatives of workers' compensation companies in Wisconsin and Minnesota and is now available on the calculator tab at SafeReturnToWork.org. Furthermore, a 90-second educational whiteboard video (Figure 3) was developed to assist with this same effort—educating farm owners or managers—the script was reviewed by 3 representatives of workers' compensation insurance companies in Wisconsin and Minnesota. This Web-based video had been disseminated to farmers via email and Facebook posts by the Professional Dairy Producers of Wisconsin, an original key informant of this project.

Table 1. Summary of focus groups (N=35).

Focus group number	Language spoken during focus group	Males, n (%)	Females, n (%)	Location	Wisconsin region
1	Spanish	5 (14)	0 (0)	Dairy farm	Central
2	English	8 (23)	0 (0)	Dairy farm	Western
3	Spanish	7 (20)	0 (0)	Dairy farm	Western
4	English	4 (11)	2 (6)	Dairy farm	Central
5	Spanish	8 (23)	1 (3)	Dairy farm	South Central

Figure 3. Screenshots of the 90-second whiteboard video. (Source: created by authors).

Discussion

Principal Findings

The owner-operator of an agricultural operation can often describe the entirety of his business to a clinician treating him or her for an injury. The owner-operator can incorporate and test his or her own injury-imposed limitations in the process of completing the workday. However, an employee generally has less knowledge of alternative work options and does not have the autonomy to decide which job tasks to take on. The clinician must establish limitations for the injured worker that are transmitted to the workplace decision maker (owner-operator or supervisor) who interprets these limitations and decides what work is possible. All too often, the workplace decision maker, confused by the speculative limitations, declares the worker unsuitable for RTW. This was reaffirmed in the interviews with farm employers.

The availability of the prototype developed during this project may continue to open communication lines with rural practitioners. Electronically linking relevant training materials to the decisions being made during the interaction with the program will put educational content and materials at the fingertips of the user, providing clinical decision support at the point of care.

Data collection with useful and actionable clinical decision support is important, if not critical, when implementing clinical systems. Once the prototype has been further tested, we anticipate expansion of the capabilities of the program by including data-gathering capabilities, allowing clinicians to track various aspects of patients' RTW. This could include the time from injury to various stages of work activities. In future iterations, we expect to include data capture such as demographics, injury type, and progress monitoring. These will provide patients and providers with the ability to understand the success of their RTW activities and track their progress in RTW management over time.

Limitations

While this project collected a near complete compendium of tasks likely to be found on dairy and pork farms, there may be substantial variability among smaller farms that this project cannot accommodate. However, smaller farms employ fewer employees, and some are run solely by the owner and family members. Thus, they are less likely to fall within a formal RTW program overseen by an employer and guided by a health care team. The system does, however, cover the vast majority of tasks on large farms and information appropriate for the majority of workers in the dairy and pork industries. Furthermore, the participants of the study and the host sites used for ergonomic data collection were primarily in Wisconsin and may not be representative of dairy and pork operations outside of the Upper Midwest.

Analyzing the functional job demands is a task that will present some variability. The descriptions by one observer may differ to some degree from the descriptions of another. We used single individuals in these tasks and did not perform repeat analyses. We chose to work with occupational therapists who were trained in the DSI methodology in an attempt to limit this variability. The DSI training is thorough and consistent, and we anticipated that this approach would reduce variability.

Conclusions

Farmers and farmworkers are increasingly adopting new technologies in their personal and professional lives, and the agricultural sector is quickly advancing in technological sophistication, from robotic milking systems to autonomous tractors and unmanned aerial vehicles. It was anticipated that farmers and farmworkers would not only be suitable as key informants, but they would also be critical in the development of an application designed to benefit the farming industry.

This translational health informatics project has produced a prototype that could be useful to rural practitioners caring for patients injured in dairy and pork production work. We believe these practitioners, the workers they care for as patients, and the employers will benefit from the guidance provided by the program. Further development of the RTW system will be pursued with subsequent funding. A future line of research may include clinical trials of the program in Wisconsin and Minnesota comparing the case parameters (eg, duration of time loss, duration of light-duty, and duration until return to full duty) of the program to statistics collected using standard practice in returning workers to work in the dairy and pork industries.

Since there is a significant financial incentive to return injured workers to a light-duty job, limiting time loss, it is unlikely that clinicians would face barriers among farmers in adopting RTW technology. However, there is little incentive on the part of clinicians to adopt the technology as is. Without seamless integration into the clinical electronic health record workflows, it is unlikely that individual physicians would consistently leverage a system such as this at the point of care without external influence or incentive. Future research should also focus on bridging gaps that appear to exist between workers' compensation insurers and physicians in the RTW process.

Acknowledgments

The authors would like to thank Marie Fleisner from the Marshfield Clinic Research Institute's Office of Scientific Writing for assistance with editing the manuscript. The authors would also like to thank Iris Reyes and Emily Redmond for their help with data collection and project coordination and the Marshfield Clinic Research Institute's software development and usability testing teams. The authors also thank Andrea Mahnke for her work as a focus group facilitator and Yurany Ninco-Sanchez for her assistance with English-Spanish translation of written materials and serving as an interpreter during focus groups.

Funding for this study was made possible by the cooperative agreement award U540H010170 for the Centers for Disease Control and Prevention/National Institute for Occupational Safety and Health, the National Farm Medicine Center, and the Marshfield Clinic Research Institute.

Conflicts of Interest

None declared.

References

1. Smith GS, Wellman HM, Sorock GS, Warner M, Courtney TK, Pransky GS, et al. Injuries at work in the US adult population: contributions to the total injury burden. Am J Public Health 2005 Jul;95(7):1213-1219. [doi: 10.2105/AJPH.2004.049338] [Medline: 15983273]
2. Leigh JP, Markowitz SB, Fahs M, Shin C, Landrigan PJ. Occupational injury and illness in the United States. Estimates of costs, morbidity, and mortality. Arch Intern Med 1997 Jul 28;157(14):1557-1568. [Medline: 9236557]

3. US Department of Labor. Occupational Safety and Health Administration. Injury & Illness Recordkeeping Forms - 300, 300A, 301 URL: https://www.osha.gov/recordkeeping/RKforms.html [accessed 2018-11-30] [WebCite Cache ID 74JxD3Z8H]
4. Kirkhorn SR, Earle-Richardson G, Banks RJ. Ergonomic risks and musculoskeletal disorders in production agriculture: recommendations for effective research to practice. Journal of Agromedicine 2010 Jul;15(3):281-299. [doi: 10.1080/1059924X.2010.488618] [Medline: 20665313]
5. Nordstrom DL, Layde PM, Olson KA, Stueland D, Brand L, Follen MA. Incidence of farm-work-related acute injury in a defined population. Am J Ind Med 1995 Oct;28(4):551-564. [Medline: 8533795]
6. Gerberich SG, Church TR, McGovern PM, Hansen HE, Nachreiner NM, Geisser MS, et al. An epidemiological study of the magnitude and consequences of work related violence: the Minnesota Nurses' Study. Occup Environ Med 2004 Jun;61(6):495-503 [FREE Full text] [Medline: 15150388]
7. Rautiainen RH, Reynolds SJ. Mortality and morbidity in agriculture in the United States. J Agric Saf Health 2002 Aug;8(3):259-276. [Medline: 12363178]
8. Waller JA. Injuries to farmers and farm families in a dairy state. J Occup Med 1992 Apr;34(4):414-421. [Medline: 1564580]
9. Pratt DS, Marvel LH, Darrow D, Stallones L, May JJ, Jenkins P. The dangers of dairy farming: the injury experience of 600 workers followed for two years. Am J Ind Med 1992;21(5):637-650. [Medline: 1609811]
10. Douphrate DI, Rosecrance JC, Wahl G. Workers' compensation experience of Colorado agriculture workers, 2000-2004. Am J Ind Med 2006 Nov;49(11):900-910. [doi: 10.1002/ajim.20387] [Medline: 17036351]
11. Gurian-Sherman D. Union of Concerned Scientist. Cambridge, MA: UCS Publications; 2008 Apr. CAFOs Uncovered The Untold Costs of Confined Animal Feeding Operations URL: http://www.ucsusa.org/assets/documents/food_and_agriculture/cafos-uncovered.pdf [accessed 2017-10-10] [WebCite Cache ID 6u6kU5TAH]
12. Jesse E. MARKETING AND POLICY BRIEFING PAPER. University of Wisconsin-Madison: Department of Agriculture and Applied Economics, College of Agricultural and Life Sciences; 2008 Nov. Growth and Transition in Wisconsin Dairying URL: https://api.aae.wisc.edu/pubs/pdf/mpbpapers/mpb96.pdf [accessed 2018-11-30] [WebCite Cache ID 6u6lEgLNN]
13. Harrison J, Lloyd S, O'Kane T. University of Wisconsin. Overview of immigrant workers on Wisconsin dairy farms. Briefing no. 1 URL: http://www.pats.wisc.edu/pubs/98 [accessed 2017-10-10] [WebCite Cache ID 6u6lgTO0E]
14. Thilmany D. Colorado State University. Colorado's agricultural workforce dynamics URL: http://www.ext.colostate.edu/emptrans/feature0302.html [accessed 2017-10-10] [WebCite Cache ID 6u6n7l0jM]
15. Rosson P, Adcock F, Susanto D, Anderson D. The economic impacts of immigration on U.S. dairy farms. College Station, TX: AgriLife Research Texas A&M System; 2009.
16. Liebman A, Juarez-Carrillo PM, Reyes IAC, Keifer MC. Immigrant dairy workers' perceptions of health and safety on the farm in America's Heartland. Am J Ind Med 2016 Mar;59(3):227-235 [FREE Full text] [doi: 10.1002/ajim.22538] [Medline: 26523613]
17. NIOSH. Analysis of the Bureau of Labor Statistics Census of Fatal Occupational Injuries microdata. Morgantown, WV: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health 2018 (unpublished).
18. Patient PHCA. One Hundred Eleventh Congress of the United States of America, the Second Session. 2010 Jan 05. Transforming the Health Care Delivery System URL: https://www.gpo.gov/fdsys/pkg/BILLS-111hr3590enr/pdf/BILLS-111hr3590enr.pdf [accessed 2017-10-10] [WebCite Cache ID 6u6mwhXaL]
19. Pransky G, Katz JN, Benjamin K, Himmelstein J. Improving the physician role in evaluating work ability and managing disability: a survey of primary care practitioners. Disabil Rehabil 2002 Nov 10;24(16):867-874. [doi: 10.1080/09638280210142176] [Medline: 12450463]
20. Guzman J, Yassi A, Cooper JE, Khokhar J. Return to work after occupational injury. Family physicians' perspectives on soft-tissue injuries. Can Fam Physician 2002 Dec;48:1912-1919 [FREE Full text] [Medline: 12520791]
21. Special Committee on Health, Productivity,Disability Management, American College of OccupationalEnvironmental Medicine. Healthy workforce/healthy economy: the role of health, productivity, and disability management in addressing the nation's health care crisis: why an emphasis on the health of the workforce is vital to the health of the economy. J Occup Environ Med 2009 Jan;51(1):114-119. [doi: 10.1097/JOM.0b013e318195dad2] [Medline: 19136880]
22. Shaw WS, van der Windt DA, Main CJ, Loisel P, Linton SJ. Early patient screening and intervention to address individual-level occupational factors ("blue flags") in back disability. J Occup Rehabil 2009 Mar;19(1):64-80. [doi: 10.1007/s10926-008-9159-7] [Medline: 19082875]
23. Shaw W, Hong Q, Pransky G, Loisel P. A literature review describing the role of return-to-work coordinators in trial programs and interventions designed to prevent workplace disability. J Occup Rehabil 2008 Mar;18(1):2-15. [doi: 10.1007/s10926-007-9115-y] [Medline: 18080827]
24. Shaw WS, Linton SJ, Pransky G. Reducing sickness absence from work due to low back pain: how well do intervention strategies match modifiable risk factors? J Occup Rehabil 2006 Dec;16(4):591-605. [doi: 10.1007/s10926-006-9061-0] [Medline: 17086499]
25. Shaw WS, Pransky G, Patterson W, Winters T. Early disability risk factors for low back pain assessed at outpatient occupational health clinics. Spine (Phila Pa 1976) 2005 Mar 01;30(5):572-580. [Medline: 15738793]

26. Tamminga SJ, van Hezel S, de Boer AG, Frings-Dresen MH. Enhancing the Return to Work of Cancer Survivors: Development and Feasibility of the Nurse-Led eHealth Intervention Cancer@Work. JMIR Res Protoc 2016 Jun 10;5(2):e118 [FREE Full text] [doi: 10.2196/resprot.5565] [Medline: 27286819]
27. Keifer MC, Reyes I, Liebman AK, Juarez-Carrillo P. The use of audience response system technology with limited-english-proficiency, low-literacy, and vulnerable populations. J Agromedicine 2014;19(1):44-52. [doi: 10.1080/1059924X.2013.827998] [Medline: 24417531]
28. Liebman AK, Juarez-Carrillo PM, Reyes IAC, Keifer MC. Immigrant dairy workers' perceptions of health and safety on the farm in America's Heartland. Am J Ind Med 2016 Mar;59(3):227-235. [doi: 10.1002/ajim.22538] [Medline: 26523613]
29. Wainwright M, Russell A. social research Update. Using NVivo Audio-Coding: Practical, Sensorial and Epistemological Considerations URL: http://sru.soc.surrey.ac.uk/SRU60.pdf [accessed 2018-08-10] [WebCite Cache ID 71ZKMxyZR]

Abbreviations

DSI: Decision support initiative
HIPAA: Health Insurance Portability and Accountability Act of 1996
MCHS: Marshfield Clinic Health System
RTW: return to work
SRTW: SafeReturnToWork.org

Edited by G Eysenbach; submitted 22.12.17; peer-reviewed by R Kanter, IN Gomez, R Pankomera, S Gorucu; comments to author 18.03.18; revised version received 13.09.18; accepted 27.09.18; published 29.01.19.

Please cite as:
Weichelt B, Bendixsen C, Keifer M
Farm Owners and Workers as Key Informants in User-Centered Occupational Health Prototype Development: A Stakeholder-Engaged Project
J Med Internet Res 2019;21(1):e9711
URL: https://www.jmir.org/2019/1/e9711/
doi:10.2196/jmir.9711
PMID:30694202

Original Paper

The Absence of Evidence is Evidence of Non-Sense: Cross-Sectional Study on the Quality of Psoriasis-Related Videos on YouTube and Their Reception by Health Seekers

Simon M Mueller[1], MD; Pierre Jungo[1], MD; Lucian Cajacob[1], MD; Simon Schwegler[1], MD; Peter Itin[1], MD; Oliver Brandt[1], MD

Department of Dermatology, University Hospital Basel, Basel, Switzerland

Corresponding Author:
Simon M Mueller, MD
Department of Dermatology
University Hospital Basel
Petersgraben 4
Basel, 4031
Switzerland
Phone: 41 61 328 69 64
Email: simon.mueller@usb.ch

Abstract

Background: Approximately 80% of internet users access health information online and patients with chronic illnesses especially rely on internet-based resources. YouTube ranks second among the most accessed websites worldwide and hosts an increasing number of videos with medical information. However, their quality is sometimes unscientific, misleading, or even harmful.

Objective: As little is known about YouTube as a source of information on psoriasis, we aimed to investigate the quality of psoriasis-related videos and, if necessary, point out strategies for their improvement.

Methods: The quality of the 100 most viewed psoriasis-related videos was assessed using the DISCERN instrument and the Global Quality Scale (GQS) by categorizing the videos into useful, misleading, and dangerous and by evaluating the reception of the videos by users.

Results: Evaluation of the videos exhibited a total of 117,221,391 views and a total duration of 10:28 hour. The majority of clips contained anecdotal personal experiences with complementary and alternative psoriasis treatments, topical treatments, and nutrition and diets being the most frequently addressed topics. While advertisements accounted for 26.0% (26/100) of the videos, evidence-based health information amounted to only 20.0% (20/100); 32.0% (32/100) of the videos were classified as useful, 52.0% (52/100) as misleading, and 11.0% (11/100) as even dangerous. The quality of the videos evaluated by DISCERN and GQS was generally low (1.87 and 1.95, respectively, on a 1 to 5 scale with 5 being the maximum). Moreover, we found that viewers rated poor-quality videos better than higher quality videos.

Conclusions: Our in-depth study demonstrates that nearly two-thirds of the psoriasis-related videos we analyzed disseminate misleading or even dangerous content. Subjective anecdotal and unscientific content is disproportionately overrepresented and poor-quality videos are predominantly rated positively by users, while higher quality video clips receive less positive ratings. Strategies by professional dermatological organizations are urgently needed to improve the quality of information on psoriasis on YouTube and other social media.

(J Med Internet Res 2019;21(1):e11935) doi:10.2196/11935

KEYWORDS

psoriasis; YouTube videos; layperson; poor quality; misleading information; dangerous content

Introduction

Social media has become increasingly important in the context of health care [1,2], and 80% of internet users access health information online, especially patients with chronic illnesses who rely on internet-based resources [2-5]. In particular, videos are powerful tools to disseminate medical information [2,3,6]. YouTube, an open access video-sharing platform, ranks second among the most accessed websites worldwide and hosts an

increasing number of videos with medical information [7-9]. YouTube counts 5 billion visits per day [7,9] and 1 billion hours watched daily [8]. Distribution of medical information to such a huge audience offers invaluable opportunities but also dangers as the quality of unfiltered information posted is often unscientific, misleading, or even harmful [2,5,10-17]. While the role and quality of YouTube videos have already been investigated in various medical specialties, only little is known about this topic in dermatology. A descriptive analysis of 100 videos covering dermatology, sun protection, skin cancer, skin cancer awareness, and skin conditions yielded over 47 million views reflecting the high demand for dermatological information posted on YouTube [18]. Psoriasis is a hot topic in social media with psoriasis foundations and associations being among the most popular dermatology-related organizations on Facebook, Twitter, and LinkedIn [19]. The benefits of these social networks providing psoriasis patients with educational information has been reported [20]. Facebook and Instagram recently attracted public attention after removing psoriasis images that they categorized as content that may not meet community guidelines [21]. Two previous publications showed that YouTube is heavily accessed as a source of information on psoriasis [22,23]. They pointed to a dominance of privately posted videos and a lack of evidence-based medical information from trustworthy institutions. However, as little is known about the scientific quality of these videos, we sought to analyze psoriasis-related videos using the Global Quality Scale (GQS) and the DISCERN tool (note that DISCERN is not an acronym but the name of the instrument) and by classifying the videos as useful, misleading, or dangerous. Furthermore, since we hypothesized that it might be difficult for laypersons to adequately judge the quality of videos, we correlated the quality of the videos with the numbers of likes and dislikes to assess viewers' ability to recognize high- and low-quality content. In addition, we analyzed the topics covered in the videos and their license types and upload sources in order to obtain a comprehensive picture of psoriasis-related YouTube videos. In summary, the objectives of this study were as follows:

- Identify upload sources, common topics, and YouTube categories of the 100 most-viewed videos
- Investigate the quality of YouTube videos as a source of information on psoriasis by applying two different score instruments
- Correlate viewers' ratings with our quality assessments
- Point out strategies for interventions that increase the quality of psoriasis video clips and medical content in general uploaded to YouTube and other social media platforms

Methods

Data Collection

In this cross-sectional study, YouTube was searched on July 27, 2017, using the term psoriasis and the filter settings English UK (language) and United Kingdom (country). Subsequently, videos were sorted by their view count. Non-English videos or channels were excluded until the top 100 videos in English were displayed (Multimedia Appendix 1). We decided to limit our analysis to the first 100 clips since this is a common and accepted procedure when investigating YouTube videos [13,16,18,24,25]. Furthermore, in our study, the first 100 videos achieved a total of 73 million views, whereby the clip in the hundredth place only achieved about 40,000 views. This suggests that videos ranked below the first 100 have only a minor impact on the outcomes.

After collecting qualitative and quantitative data (duration, upload data, source, likes/dislikes, category, license type), overall quality of the videos was assessed by 5 experienced dermatologists using two assessment tools [26]. To optimize interrater agreement on the videos, the dermatologists attended training sessions to get familiar with quality assessments and the rating policy and criteria.

Creation of Content Categories

In a first step, topics of the video clips were collected. If a video covered more than one topic, each topic was listed separately. The content was subsequently categorized according to commonalities and by topics and/or categories discussed in two previous YouTube studies on psoriasis [22,23]. Unlike other studies that only used titles to categorize topics, we exclusively considered content for categorization, as the title often does not reflect the actual content of the clip.

Scoring and Classification of Videos

The GQS, which is based on a 5-point scale, was developed in 2012 by Singh et al [16] for the evaluation of YouTube videos and has since been applied in numerous studies. The score measures the quality and flow of information and the value of an information source for medical laymen (Multimedia Appendix 2, Table A).

The DISCERN instrument is used to measure the quality of health information about treatment choices provided in video clips [27]. Originally developed for the standardized assessment of written medical information, the DISCERN tool consists of 15 key questions and an evaluation of overall quality with which the reliability (questions 1 to 8), quality of information (questions 9 to 15) and overall quality of a publication (question 16) can be assessed by assigning 1 to 5 points per question (Multimedia Appendix 2, Table B). For both tools, the higher the total value, the higher the quality of the video clip.

In addition, videos were classified into useful, misleading, or dangerous and categorized by topic or content, presence and profession of a presenter (to be seen in the video, health or nonhealth professional), and upload sources. In case of differing assessments by the analyzing dermatologists, the corresponding video was reassessed by the principal investigator (SM).

Statistical Analyses

Descriptive statistics and Spearman correlation coefficients to calculate the number of likes and dislikes with the values of DISCERN and GQS, respectively, were performed using SPSS Statistics version 22.0 (IBM Corp). To assess the interrater reliability, the Cohen kappa coefficients and intraclass correlation coefficients were calculated.

Results

View Count, Duration, Upload Sources, Categories, and Topics

The 100 videos had a total of 117,221,391 views and a total duration of 10:28 hours (mean duration per video 6:17 [SD 6:39] minutes). The two most viewed videos accounted for 93,736,280 views (79.96% of the total) and were pharmaceutical advertising with a Creative Commons license and likes, dislikes, and comment functions disabled. This kind of license authorizes users not only to download the video but also to use the entire clip or parts thereof for their own video clip productions. The other 98 videos had a Standard YouTube License, which allows the use of the clips only after prior permission of the author.

The majority (65/100, 65.0%) of the videos were uploaded from the United States. The most frequent category was People & Blogs (36/100, 36.0%), a diverse mix of content and the most famous category on YouTube [28], while 27.0% (27/100) appeared in the category Education, 21.0% (21/100) in Howto & Style and only 7.0% (7/100) in Science and Technology.

Most videos contained anecdotal, personal, unscientific, or commercial information on psoriasis (Figure 1). While pharmaceutical and nonpharmaceutical advertisements made up 26.0% (26/100) of the video clips, evidence-based health information posted by professional individuals and institutions amounted to only 20.0% (20/100). Complementary and alternative psoriasis treatments (41/100, 41.0%), topical treatments (39/100, 39.0%), and nutrition/diets (25/100, 25.0%) were the most often addressed topics (Figure 2). Video clips were uploaded between September 2007 and June 2016, with numbers increasing from 2012 on (Figure 3). In some videos, the comment function was disabled and the YouTube statistics were not declared.

Figure 1. Distribution of topics, information providers and upload sources (multiple categories may apply to one video); *including websites from psoriasis associations.

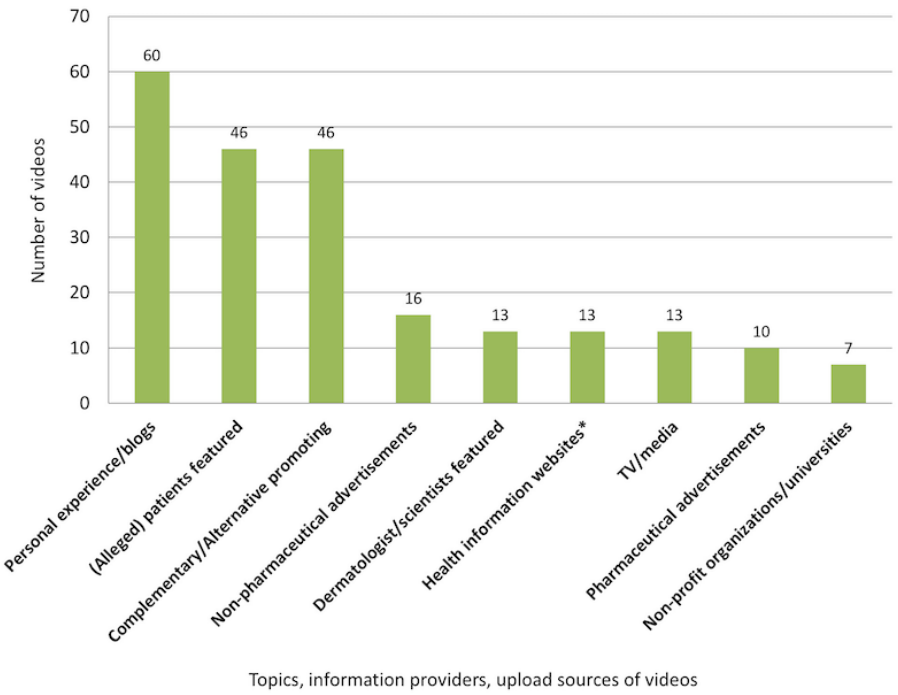


Figure 2. Topics presented in the videos (Note: a video clip can cover more than one topic; *alternative treatment includes complementary treatment; UVR: ultraviolet radiation).

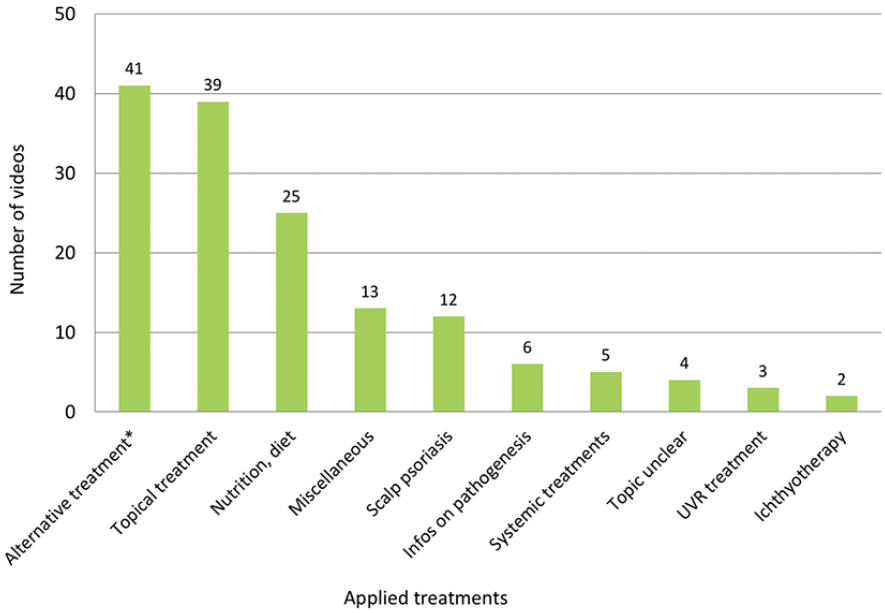


Figure 3. Distribution of the uploaded videos over the period 2007-2016 (n=100).

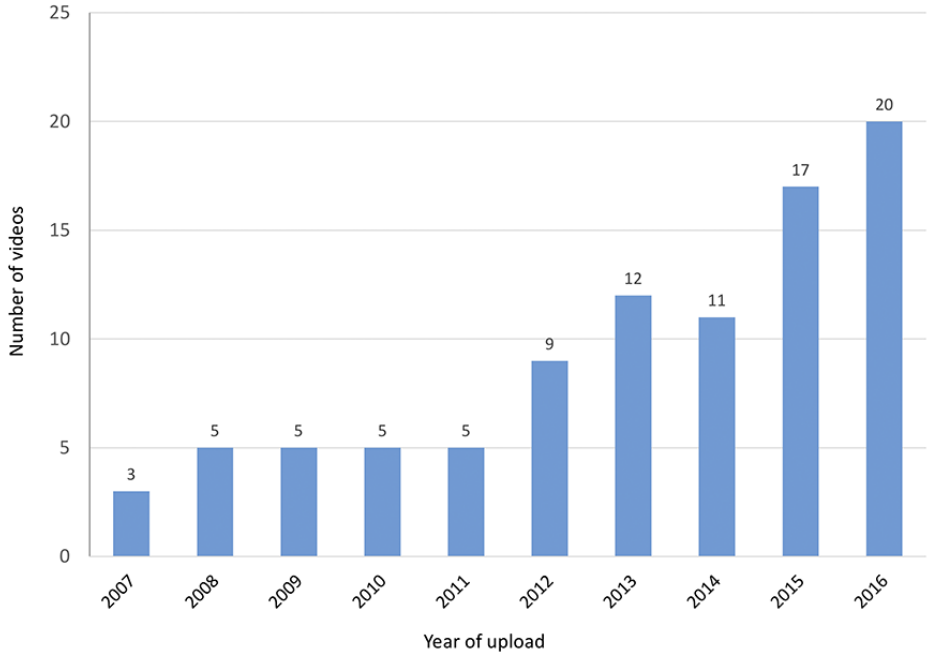
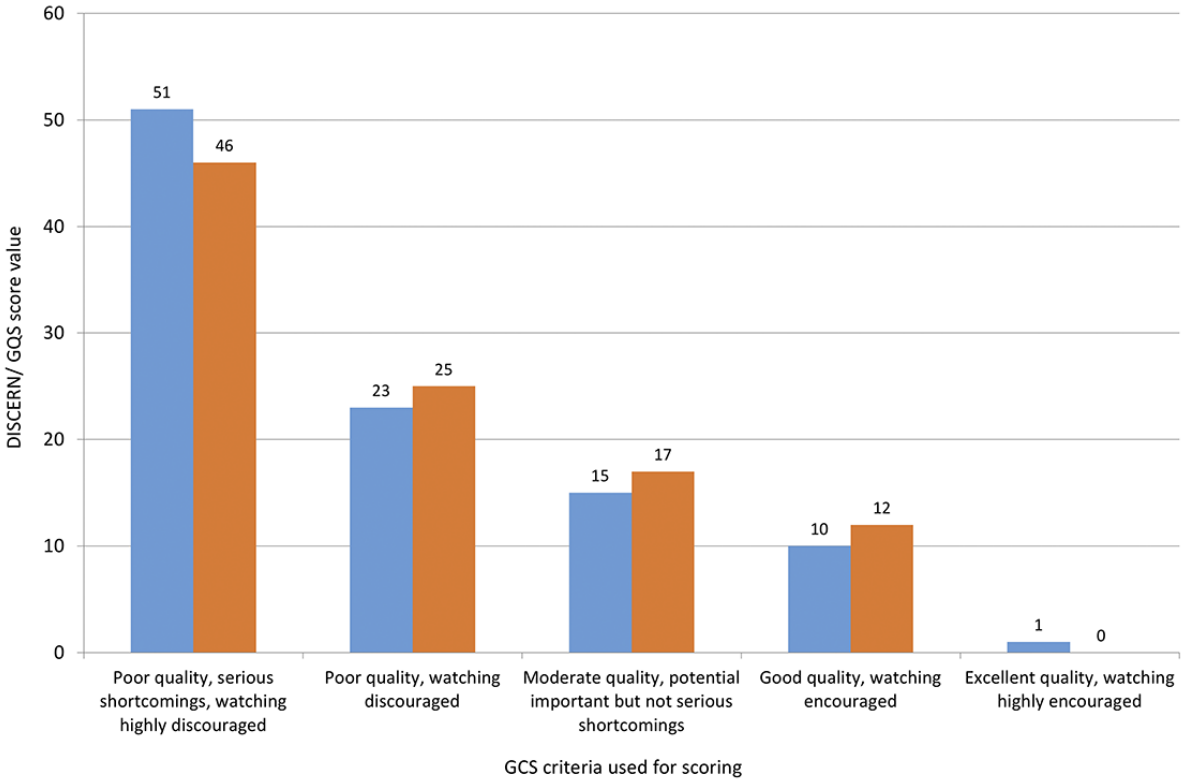


Figure 4. Comparison of quality assessments of the videos (n=100) performed with the DISCERN instrument (blue bars) and the Global Quality Scale (orange bars).



Quality Assessments and Correlation With Likes and Dislikes

Of the videos, 32.0% (32/100) were classified as useful and 63.0% (63/100) as misleading—of these, 17.5% (11/63) were even considered dangerous because of potential mechanical or chemical injury or harmful recommendations regarding sun exposure or diets; 5% (5/100) of the videos were neither useful nor misleading.

With a value of .74, the kappa statistic revealed a good level of agreement among the raters.

In terms of the view count, we excluded the 2 most viewed videos from further analyses as they were pharmaceutical advertisements, accounting for 79.96% (93,736,280/117,221,391) of all views. The misleading videos (63/100) garnered 18,387,077 views including 4,611,126 views of videos with potentially dangerous content (11/63). Useful videos (32/100) had 5,098,034 views resulting in a ratio of 3.61 (18,387,077:5,098,034) misleading to useful videos.

The ratings using the DISCERN and GQS scores were consistent, yielding the categorizations shown in Figure 4. The quality of the videos, expressed by the mean overall DISCERN and GQS rating scores, was generally low (1.87 [SD 1.07] and 1.95 [SD 1.06], respectively) on a 1 to 5 scale with 5 being the maximum. Detailed analysis of the DISCERN values revealed that the major shortcomings were lack of information about the evidence and source of the posted information, areas of uncertainty and risks of the praised therapy, and missing recommendations for shared decision making or links to additional sources of information (see Multimedia Appendix 3).

The intraclass correlation coefficients calculated for the DISCERN and GQS were .81 and .78, respectively, indicating a high level of agreement between the assessors.

The videos received 113,147 likes and 9260 dislikes yielding a like to dislike ratio of 12.4 (113,147:9260). In 7 videos, the like/dislike function was disabled. When correlating the viewer ratings with our quality assessments, we found a negative correlation between the number of likes and the DISCERN mean values (Spearman correlation coefficient ρ=–0.23, P=.24) and a negative correlation between the number of dislikes and the DISCERN and GQS mean values (ρ=–0.34, P=.001, and ρ=–0.37, P<.001, respectively), meaning that viewers rated poor quality better than higher quality videos.

Discussion

Principal Findings

Psoriasis patients are avid users of social media, including YouTube, as a source of information on their disease [19-23,29]. However, little is known about the upload sources, topics, and particularly the scientific quality of these YouTube videos. Moreover, it is unknown whether viewer ratings correlate with the quality of the medical information posted.

This study found that the majority of video clips contained anecdotal personal experiences mainly addressing topics such as alternative treatment options for psoriasis and putative benefits of diets. Alarmingly, more than half of the videos spread misleading and about 1 in 10 even dangerous information and

recommendations. Furthermore, the quality of the video clips was rather low, and the fact that viewers rated poor quality better than higher quality videos indicates that the majority of health seekers are not capable to recognize low quality medical information in videos as such.

Misleading Information

We found that nearly two-thirds of the top 100 psoriasis-related YouTube videos disseminate misleading information. While this proportion of misleading content is in line with previous nondermatological studies [14,15,17,30,31], no direct comparisons in terms of psoriasis-related studies are currently available. Qi et al [23] have analyzed 47 psoriasis videos with a total of 2 million views but only distinguished between useful (18 videos) and misleading (10 videos), which makes a direct comparison with our results impossible. Lenczowski et al [22], on the other hand, did not judge the quality of the psoriasis videos they investigated.

It is unclear why the two pharmaceutical company videos account for almost 80% (93,736,280) of the more than 117 million visitors to the top 100. As the statistical information provided by YouTube did not allow us to determine whether the videos accessed were actually or completely viewed and when they were left, the significance of this figure is limited. It is, however, conceivable that at least some of the visits were due to the Creative Commons License, which, as mentioned above, allows the use of YouTube videos for personal video clip productions.

Potentially Dangerous Content

A total of 11% of the videos we analyzed contained potentially dangerous content. For example, psoriasis patients were encouraged to remove their plaques using a knife blade, glue, Brazilian waxing, and apple cider vinegar. In addition, sunbathing without reference to sun protection, unnecessary diets (eg, avoidance of dairy or gluten), and the use of the one and only miracle cure were praised in such videos. This advice was frequently posted by patients reporting a personal negative long-term experience with conventional medicine who eventually found salvation in alternative treatments. It has been reported that patients with moderate to severe psoriasis are more apt to rely on psoriasis user–generated content in social media than their counterparts suffering from milder forms [29]. Therefore, it can be assumed that patients with more severe forms may be more prone to follow dubious advice and thus have an increased risk for undertreatment of skin and joint inflammations favoring the progressive psoriatic march.

Low Percentage of Good or Excellent Quality Videos

Interestingly, the main topics published in psoriasis-related videos revolve around complementary and alternative psoriasis treatments, (homemade) current therapies, and nutrition and diet topics that allow patients to take measures to improve their skin condition without consulting a physician or health care professional. According to the results of the GQS and DISCERN tools, only 11% and 12% of the videos, respectively, were of good or excellent quality with unbiased, evidence-based or at least science-based information. This and our analyses of the uploaded sources indicate, in accordance with Lenczowski et al [22], that health care organizations, universities, and dermatologists are clearly underrepresented on YouTube in the context of psoriasis. The issue of a lack in high-quality information seems to be exacerbated by the negative correlation we found between the quality and the number of likes: high-quality videos are not as popular as low-quality videos. This finding is in line with Qi et al [23], who reported that useful psoriasis videos had fewer likes than misleading ones. The trend of nonuseful videos being more popular than useful ones has also been reported by nondermatological studies [11,13,31]. It remains unclear why viewers appear to like low-quality videos more than high-quality ones, and we can only speculate about the reasons. It is conceivable that (1) they just do not recognize high quality, (2) high-quality videos are too complex and less entertaining, or (3) viewers are intentionally looking for unconventional content diverging from established medical recommendations. The latter possibility may be supported by the observation of Lenczowski et al [22], which suggests that unconventional videos receive more views and likes than traditional medical videos. Remarkably, not only did a high number of likes correlate with low quality, a low number of positive ratings correlated with high-quality video clips. Exploration of viewer comments could help to elucidate the relationship between the likes and dislikes and the quality of the content.

These findings raise question about why certain viewers are so drawn to low-quality videos and how best to deal with this phenomenon to bring about change. To answer this, it is first of all helpful to gain knowledge of the characteristics of the typical psoriasis health seeker [22,32]. Previous studies demonstrated that health seekers in general have limited skills in searching and evaluating medical content on the internet and rarely call up results that appear beyond the second results page [33,34]. Websites without commercial advertising using medical terms enjoy more trust among the majority of users and are not left as quickly as those that do not meet these criteria [33,35]. Moreover, the majority of individuals searching the internet for medical information feel confident when advice matches what they already know or think they know about the subject and when similar information about it is available on more than one site [32,35]. Furthermore, we found that more than one-third of the videos we analyzed were uploaded to the popular YouTube category People & Blogs, which allows conclusions to be drawn about the claims and ideas of those who address the public with their supposed knowledge. From these findings, some of the necessary steps toward a better information policy for psoriasis patients in the social media can be derived.

Possible Interventions

First, it is urgently necessary that dermatology associations, psoriasis self-help organizations, etc, provide medically accurate, high-quality, and easy-to-understand information—including videos—for laymen dealing not only with pathophysiology, clinical manifestations, and evidence-based therapeutic options for psoriasis but also with non–evidence-based treatments and their inefficacy and potential hazards.

Second, it is important to keep in mind that YouTube videos reach a large audience, a fact that the World Health Organization

has recognized; it explicitly recommends the use of this platform in its strategic communications framework [36]. In our analysis, YouTube provided more than 5 million viewers with partly useful information on psoriasis. However, 3.6 times more viewers watched videos with misleading information indicating that YouTube may be a double-edged sword. Most viewers are likely to watch several videos including both misleading and useful ones. As it is conceivable that misconceptions may be corrected by useful information, it is crucial that the number of videos on psoriasis posted by professional health organizations is sufficient to neutralize the misleading ones. Quality assurance measures are essential to achieve this.

To our knowledge, this is the first study to include the DISCERN tool in the context of dermatology-related YouTube videos. The mean score gained from the overall rating of the DISCERN tool (question number 16) was very similar to the GQS score (both approximately 1.9). However, as the DISCERN scoring system provides important additional information, we believe that this tool is better suited for quality assessment of videos and should therefore also be used to evaluate films from professional health organizations before they are posted.

Third, as the majority of health seekers rely on the first 10 results from the search engine, it is important to ensure that websites, videos, etc, of professional health care providers appear on the first 2 result pages [33]. To achieve this, search engine optimization can be performed by using as many keywords that health seekers regularly use in their research as possible in domains and meta elements (metatags), headlines, and in the bodies of the text [34].

Furthermore, cooperation between social media and search engine providers and dermatology associations and psoriasis self-help organizations would enable the positioning of medically accurate information in a prominent location on the results page and easier access to corresponding websites and videos. Such an approach had already proved successful in 2003 during the SARS epidemic, when the internet company Google ensured that the World Health Organization and Centers of Disease Control and Prevention websites were displayed at the top of the first results page in its search engine [32].

Finally, one could also think about subjecting websites and video clips with medical content to standardized quality control or setting minimum standards through government regulations. However, due to the amount of media posted daily and the philosophy of the YouTube platform, this seems hardly feasible.

Strengths and Limitations

Despite their unmistakable strengths, such as the comprehensive analyses of a high number of videos and the application of two different scoring tools (GQS and DISCERN), there are some limitations to our study. Although we performed comprehensive analyses, we neither evaluated the comments posted by viewers nor did we investigate potential associations between the duration of YouTube videos and their quality, the number of likes and dislikes they received, etc, which might have allowed us, among other things, to make statements about preferences of viewers and optimal durations of video clips.

Conclusion

Our study demonstrates and confirms the results of others that the vast majority of psoriasis videos presented on YouTube contain misleading and sometimes potentially harmful information about this disease. Moreover, our findings suggest that a large number of users looking for clips on psoriasis on YouTube are not only unable to distinguish between medically accurate and inaccurate information but even tend to rate videos of inferior quality better than videos of higher quality.

According to our findings and in agreement with the five previous dermatological YouTube studies [18,22,23], it is crucial that dermatology associations and health facilities identify the motives of the users for this behavior. This would enable them, together with operators of social media platforms and state institutions, to develop strategies aimed at improving the quality of the information provided on YouTube and other social media platforms. The evidence-based material should be created in such a way that it can easily be found by search engines and appear on the first 2 results pages.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Links to YouTube videos.

[XLSX File (Microsoft Excel File), 19KB - jmir_v21i1e11935_app1.xlsx]

Multimedia Appendix 2

Scores used for quality evaluation of video clips.

[PDF File (Adobe PDF File), 27KB - jmir_v21i1e11935_app2.pdf]

Multimedia Appendix 3

Rating of the 15 items of the DISCERN tool on a scale from 1 to 5 (with 5 being the maximum).

[PDF File (Adobe PDF File), 155KB - jmir_v21i1e11935_app3.pdf]

References

1. Samuel N, Alotaibi NM, Lozano AM. YouTube as a source of information on neurosurgery. World Neurosurg 2017 Sep;105:394-398. [doi: 10.1016/j.wneu.2017.05.111] [Medline: 28599904]
2. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194. [doi: 10.1177/1460458213512220] [Medline: 24670899]
3. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
4. Fat MJL, Doja A, Barrowman N, Sell E. YouTube videos as a teaching tool and patient resource for infantile spasms. J Child Neurol 2011 Jul;26(7):804-809. [doi: 10.1177/0883073811402345] [Medline: 21551373]
5. Hayanga AJ, Kaiser HE. Medical information on YouTube. JAMA 2008 Mar 26;299(12):1424-1425. [doi: 10.1001/jama.299.12.1424-b] [Medline: 18364483]
6. Smith AA. YouTube your science. Nature 2018 Dec;556(7701):397-398. [doi: 10.1038/d41586-018-04606-2] [Medline: 29666498]
7. Alexa. Alexa website traffic analysis URL: https://www.alexa.com/siteinfo/youtube.com [WebCite Cache ID 72kjA3oqq]
8. YouTube. YouTube by the numbers URL: https://www.youtube.com/yt/about/press/ [WebCite Cache ID 73sXWrUYR]
9. 36 mind-blowing YouTube facts, figures and statistics—2017 (re-post).: Videonitch URL: http://videonitch.com/2017/12/13/36-mind-blowing-youtube-facts-figures-statistics-2017-re-post/ [accessed 2018-11-12] [WebCite Cache ID 72kmba5Xx]
10. Syed-Abdul S, Fernandez-Luque L, Jian W, Li Y, Crain S, Hsu M, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
11. Kumar N, Pandey A, Venkatraman A, Garg N. Are video sharing web sites a useful source of information on hypertension? J Am Soc Hypertens 2014 Jul;8(7):481-490. [doi: 10.1016/j.jash.2014.05.001] [Medline: 25064770]
12. Delli K, Livas C, Vissink A, Spijkervet FK. Is YouTube useful as a source of information for Sjögren's syndrome? Oral Dis 2016 Apr;22(3):196-201. [doi: 10.1111/odi.12404] [Medline: 26602325]
13. Garg N, Venkatraman A, Pandey A, Kumar N. YouTube as a source of information on dialysis: a content analysis. Nephrology (Carlton) 2015 May;20(5):315-320. [doi: 10.1111/nep.12397] [Medline: 25641264]
14. Biggs TC, Bird JH, Harries PG, Salib RJ. YouTube as a source of information on rhinosinusitis: the good, the bad and the ugly. J Laryngol Otol 2013 Aug;127(8):749-754. [doi: 10.1017/S0022215113001473] [Medline: 23866821]
15. Hegarty E, Campbell C, Grammatopoulos E, DiBiase AT, Sherriff M, Cobourne MT. YouTube as an information resource for orthognathic surgery. J Orthod 2017 Jun;44(2):90-96. [doi: 10.1080/14653125.2017.1319010] [Medline: 28463076]
16. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis—a wakeup call? J Rheumatol 2012 May;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
17. Abedin T, Ahmed S, Al Mamun M, Ahmed SW, Newaz S, Rumana N, et al. YouTube as a source of useful information on diabetes foot care. Diabetes Res Clin Pract 2015 Oct;110(1):e1-e4. [doi: 10.1016/j.diabres.2015.08.003] [Medline: 26303266]
18. Boyers LN, Quest T, Karimkhani C, Connett J, Dellavalle RP. Dermatology on YouTube. Dermatol Online J 2014 Jun 15;20(6):1 [FREE Full text] [Medline: 24945641]
19. Amir M, Sampson BP, Endly D, Tamai JM, Henley J, Brewer AC, et al. Social networking sites: emerging and essential tools for communication in dermatology. JAMA Dermatol 2014 Jan;150(1):56-60. [doi: 10.1001/jamadermatol.2013.6340] [Medline: 24196212]
20. Idriss SZ, Kvedar JC, Watson AJ. The role of online support communities: benefits of expanded social networks to patients with psoriasis. Arch Dermatol 2009 Jan;145(1):46-51. [doi: 10.1001/archdermatol.2008.529] [Medline: 19153342]
21. Stafford N. Dermatologists criticise Facebook and Instagram for removing psoriasis images. BMJ 2018 Dec 09;360:k1137. [Medline: 29523578]
22. Lenczowski E, Dahiya M. Psoriasis and the digital landscape: YouTube as an information source for patients and medical professionals. J Clin Aesthet Dermatol 2018 Mar;11(3):36-38 [FREE Full text] [Medline: 29606999]
23. Qi J, Trang T, Doong J, Kang S, Chien AL. Misinformation is prevalent in psoriasis-related YouTube videos. Dermatol Online J 2016 Nov 15;22(11):1. [Medline: 28329562]
24. Bora K, Das D, Barman B, Borah P. Are internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015-16 Zika virus pandemic. Pathog Glob Health 2018 Aug 29:1-9. [doi: 10.1080/20477724.2018.1507784] [Medline: 30156974]
25. Nagpal SJS, Karimianpour A, Mukhija D, Mohan D, Brateanu A. YouTube videos as a source of medical information during the Ebola hemorrhagic fever epidemic. Springerplus 2015;4:457 [FREE Full text] [doi: 10.1186/s40064-015-1251-9] [Medline: 26322263]
26. Drozd B, Couvillon E, Suarez A. Medical YouTube videos and methods of evaluation: literature review. JMIR Med Educ 2018 Feb 12;4(1):e3 [FREE Full text] [doi: 10.2196/mededu.8527] [Medline: 29434018]

27. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [Medline: 10396471]
28. Marshall C. Tubular Insights. 2015. How many views does a YouTube video get? Average views by category URL: http://tubularinsights.com/average-youtube-views/ [WebCite Cache ID 72kv1i5Js]
29. Steakley L. A look at social-media use among psoriasis patients. 2012. URL: https://scopeblog.stanford.edu/2012/12/21/a-look-at-social-media-use-among-psoriasis-patients/ [accessed 2018-06-13] [WebCite Cache ID 72kqfjjbM]
30. Donzelli G, Palomba G, Federigi I, Aquino F, Cioni L, Verani M, et al. Misinformation on vaccination: a quantitative analysis of YouTube videos. Hum Vaccin Immunother 2018 Jul 03;14(7):1654-1659. [doi: 10.1080/21645515.2018.1454572] [Medline: 29553872]
31. Lee JS, Seo HS, Hong TH. YouTube as a source of patient information on gallstone disease. World J Gastroenterol 2014 Apr 14;20(14):4066-4070 [FREE Full text] [doi: 10.3748/wjg.v20.i14.4066] [Medline: 24744597]
32. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [Medline: 15667044]
33. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]
34. Hansen DL, Derry HA, Resnick PJ, Richardson CR. Adolescents searching for health information on the Internet: an observational study. J Med Internet Res 2003 Oct 17;5(4):e25 [FREE Full text] [doi: 10.2196/jmir.5.4.e25] [Medline: 14713653]
35. Fox S, Rainie L. Vital decisions: a Pew internet health report. Washington: Pew Internet and American Life Project; 2002 May 22. URL: http://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2002/PIP_Vital_Decisions_May2002.pdf.pdf [accessed 2018-11-12] [WebCite Cache ID 73sYqJLUV]
36. WHO Strategic Communications Framework for effective communications. Geneva: World Health Organization URL: http://www.who.int/mediacentre/communication-framework.pdf [accessed 2018-11-12] [WebCite Cache ID 73sZDfi3b]

Abbreviations

GQS: Global Quality Score
UVR: ultraviolet radiatio

Edited by T Kool; submitted 21.08.18; peer-reviewed by K Chalil Madathil, C Basch; comments to author 20.09.18; revised version received 28.09.18; accepted 08.10.18; published 16.01.19.

Please cite as:
Mueller SM, Jungo P, Cajacob L, Schwegler S, Itin P, Brandt O
The Absence of Evidence is Evidence of Non-Sense: Cross-Sectional Study on the Quality of Psoriasis-Related Videos on YouTube and Their Reception by Health Seekers
J Med Internet Res 2019;21(1):e11935
URL: http://www.jmir.org/2019/1/e11935/
doi:10.2196/11935
PMID:30664460

Original Paper

Consumer Health Search on the Web: Study of Web Page Understandability and Its Integration in Ranking Algorithms

Joao Palotti[1,2], MSc; Guido Zuccon[3], PhD; Allan Hanbury[2,4], PhD

[1]Qatar Computing Research Institute, Doha, Qatar

[2]Institute for Information Systems Engineering, Technische Universität Wien, Vienna, Austria

[3]University of Queensland, Brisbane, Australia

[4]Complexity Science Hub Vienna, Vienna, Austria

Corresponding Author:
Allan Hanbury, PhD
Institute for Information Systems Engineering
Technische Universität Wien
Favoritenstraße 9-11/194 04
Vienna, 1040
Austria
Phone: 43 158801188310
Email: allan.hanbury@tuwien.ac.at

Abstract

Background: Understandability plays a key role in ensuring that people accessing health information are capable of gaining insights that can assist them with their health concerns and choices. The access to unclear or misleading information has been shown to negatively impact the health decisions of the general public.

Objective: The aim of this study was to investigate methods to estimate the understandability of health Web pages and use these to improve the retrieval of information for people seeking health advice on the Web.

Methods: Our investigation considered methods to automatically estimate the understandability of health information in Web pages, and it provided a thorough evaluation of these methods using human assessments as well as an analysis of preprocessing factors affecting understandability estimations and associated pitfalls. Furthermore, lessons learned for estimating Web page understandability were applied to the construction of retrieval methods, with specific attention to retrieving information understandable by the general public.

Results: We found that machine learning techniques were more suitable to estimate health Web page understandability than traditional readability formulae, which are often used as guidelines and benchmark by health information providers on the Web (larger difference found for Pearson correlation of .602 using gradient boosting regressor compared with .438 using Simple Measure of Gobbledygook Index with the Conference and Labs of the Evaluation Forum eHealth 2015 collection).

Conclusions: The findings reported in this paper are important for specialized search services tailored to support the general public in seeking health advice on the Web, as they document and empirically validate state-of-the-art techniques and settings for this domain application.

(J Med Internet Res 2019;21(1):e10986) doi:10.2196/10986

KEYWORDS

readability; literacy; comprehension; patients; machine learning

Introduction

Background

Search engines are concerned with retrieving relevant information to support a user's information-seeking task. Commonly, signals about the topicality or aboutness of a piece of information with respect to a query are used to estimate relevance, with other relevance dimensions such as understandability and trustworthiness [1] being relegated to a secondary position or completely neglected. Although this might be a minor problem for many information-seeking tasks, there are some specific tasks in which dimensions other than topicality have an important role in the information seeking and

decision-making process. The seeking of health information and advice on the Web by the general public is one such task.

A key problem when searching the Web for health information is that this can be too technical, unreliable, generally misleading, and can lead to unfounded escalations and poor decisions [2-4]. Where correct information exists, it can be hard to find and digest among the noise, spam, technicalities, and irrelevant information. In *high-stakes search tasks* such as this, access to poor information can lead to poor decisions, which ultimately can have a significant impact on our health and well-being [4,5]. In this study, we are specifically interested in the understandability of health information retrieved by search engines and in improving search results to favor information understandable by the general public. We leave addressing reliability and trustworthiness of the retrieved information to future work; however, this can be achieved by extending the framework we investigate here.

The use of general purpose Web search engines such as Google, Bing, and Baidu for seeking health advice has been largely analyzed, questioned, and criticized [6-11], despite the commendable efforts these services have put into providing increasingly better health information, for example, the Google Health Cards [12].

Ad hoc solutions to support the general public in searching and accessing health information on the Web have been implemented, typically supported by government initiatives or medical practitioner associations, for example, HealthOnNet.org (HON [13]) and HealthDirect.gov.au, among others. These solutions aim to provide *better* health information to the general public. For example, HON's mission statement is "to guide Internet users to reliable, understandable, accessible and trustworthy sources of medical and health information." On the contrary, do the solutions that these services currently employ actually provide this type of information to the health-seeking general public?

As an illustrative example, we analyzed the top 10 search results retrieved by HON on October 01, 2017 in answer to 300 health search queries generated by regular health consumers in health forums. These queries are part of the Conference and Labs of the Evaluation Forum (CLEF) 2016 electronic health (eHealth) collection [14], which is extensively used in this paper. The understandability score of the retrieved pages was estimated with the most effective readability formula (RF) and preprocessing settings analyzed in this paper (low scores correspond to easy to understand Web pages). Figure 1 reports the cumulative distribution of understandability scores for these search results (note, we did not assess their topical relevance here). Dale-Chall Index (DCI) measures the years of schooling required to understand a document. The average US resident reads at or below an 8th grade level [15-18], which is the level suggested by the American National Institutes of Health for health information on the Web [19]. We also report the scores for the *optimal* search results (Oracle), as found in CLEF 2016 (relevant results that have the highest understandability scores), along with the scores for the baseline method (Best Match 25 [BM25]) and our best retrieval method, eXtreme Gradient Boosting (XGB). The results clearly indicate that despite solutions such as HON being explicitly aimed at supporting access to high-quality health information that can aid the user to take well-informed health decisions, they often fail to direct the users to information they can understand.

In this paper, we aim to establish methods and best practice for developing search engines that retrieve *relevant and understandable* health advice from the Web. The overall contributions of this paper can be summarized as:

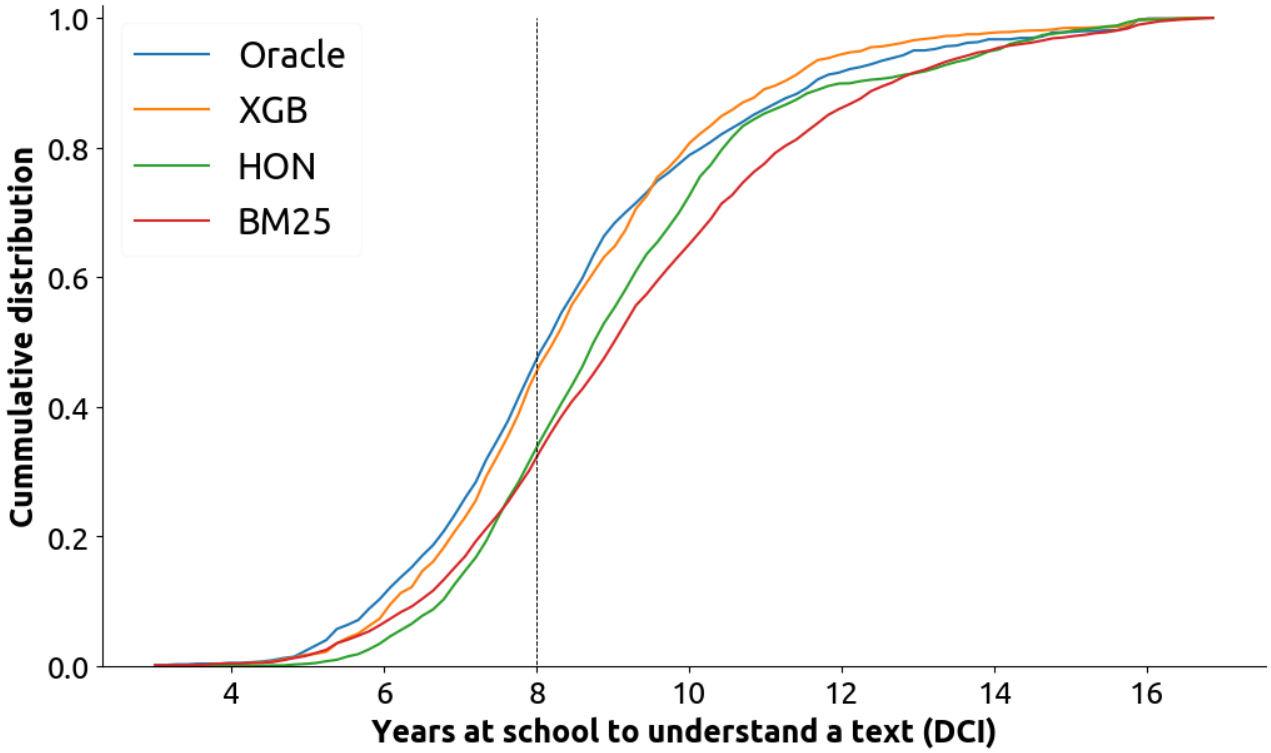
1. We propose and investigate methods for the estimation of the understandability of health information in Web pages: a large number of medically focused features are grouped in categories and their contribution to the understandability estimation task is carefully measured.
2. We further study the influence of HTML processing methods on these estimations and their pitfalls, extending our previous work that has shown how this often-ignored aspect greatly impacts effectiveness [20].
3. We further investigate how understandability estimations can be integrated into retrieval methods to enhance the quality of the retrieved health information, with particular attention to its understandability by the general public. New models are explored in this paper, also extending our previous work [21].

This paper makes concrete contributions to practice, as it informs health search engines specifically tailored to the general public (eg, the HON or HealthDirect services referred to above) about the best methods they should adopt. These are novel and significant contributions as no previous work has systematically analyzed the influence of the components in this study—we show that these greatly influence retrieval effectiveness and, thus, delivery of relevant and understandable health advice.

Related Work

Understandability refers to the ease of comprehension of the information presented to a user. In other words, health information is understandable "when consumers of diverse backgrounds and varying levels of health literacy can process and explain key messages" [22]. Often the terms understandability and readability are used interchangeably: we use readability to refer to formulae that estimate how easy it is to understand a text, usually based on its words and sentences. We use understandability to refer to the broader concept of ease of understanding: this is affected by text readability (as increasing readability tends to improve understanding) but might also be influenced by how legible a text is and its layout, including, for example, the use of images to explain difficult concepts.

Figure 1. Cumulative distribution of Dale-Chall Index (DCI) of search results. DCI measures the years of schooling required to understand a document. The dashed line is the 8th grade level which is the reading level of an average US resident. The distribution for HealthOnNet (HON) is similar to that of the baseline used in this paper (Best Match 25 [BM25]). Our best method (eXtreme Gradient Boosting [XGB]) reranks documents to provide more understandable results; its distribution is similar to that of an oracle system.



There is a large body of literature that has examined the understandability of Web health content when the information seeker is a member of the general public. For example, Becker reported that the majority of health websites are not well designed for the elderly [23], whereas Stossel et al found that health education material on the Web is not written at an adequate reading level [18]. Zheng and Yu have reported on the readability of electronic health records compared with Wikipedia pages related to diabetes and found that readability measures often do not align with user ratings of readability [24]. A common finding of these studies is that, in general, health content available on Web pages is often hard to understand by the general public; this includes content that is retrieved in top-ranked positions by current commercial search engines [6-11].

Previous linguistics and information retrieval research has attempted to devise computational methods for the automatic estimation of text readability and understandability, and for the inclusion of these within search methods or their evaluation. Computational approaches to understandability estimations include (1) *RF*, which generally exploit word surface characteristics of the text, (2) *machine learning* approaches, and (3) matching with specialized *dictionaries or terminologies*, often compiled with information about understandability difficulty.

Measures such as Coleman-Liau Index (CLI) [25], DCI [26], and Flesch Reading Ease (FRE) [27] belong to the first category. These measures generally rely on surface-level characteristics of text such as characters, syllables, and word counts [28]. Although these measures have been widely used in studies investigating the understandability of health content retrieved by search engines [6-11,18,23]), our preliminary work found that these measures are heavily affected by the methods used to extract text from the HTML source [20]. We were able to identify specific settings of an HTML preprocessing pipeline that provided consistent estimates, but because of the lack of human assessments, we were not able to investigate how well each HTML preprocessing pipeline correlated with human assessments. In this paper, we revisited and extended this work in more detail, as we further investigated this problem by comparing the effect of HTML preprocessing on text understandability estimations in light of explicit human assessments.

The use of machine learning to estimate understandability forms an alternative approach. Earlier research explored the use of statistical natural language processing and language modeling [29-31] as well as linguistic factors such as syntactic features or lexical cohesion [32]. Although we replicated here many of the features devised in these works, they focus on estimating readability of general English documents rather than medical ones. In the medical domain, Zeng et al explored features such as word frequency in different medical corpora to estimate concept familiarity, which prompted the construction of the consumer health vocabulary (CHV) [33-35].

The actual use of CHV or other terminologies such as the Medical Subject Headings (MeSH) belongs to the third category of approaches. The CHV is a prominent medical vocabulary dedicated to mapping layperson vocabulary to technical terms [34]. It attributes a score for each of its concepts with respect to their difficulty, with lower or higher scores for harder or

easier concepts. Researchers have evaluated CHV in tasks such as document analysis [36] and medical expertise prediction [37]. The hierarchy of MeSH was previously used in the literature to identify difficult concepts, assuming that a concept deep in the hierarchy is more difficult than a shallow one [38]. Other approaches combined vocabularies with word surface characteristics and syntactic features, such as part of speech (POS), into a unique readability measure [39].

In this study, we investigated approaches to estimate understandability from each of these categories, including measure the influence of HTML preprocessing on automatic understandability methods and establish best practices.

Some previous works have attempted to use understandability estimations for improving search results in consumer health search as well as methods to evaluate retrieval systems that do account for understandability along with topical relevance. Palotti et al have used learning to rank with standard retrieval features along with features based on RF and medical lexical aspects to determine understandability [21]. Van Doorn et al have shown that learning a set of rankers that provide trade-offs across a number of relevance criteria, including readability or understandability, increases overall system effectiveness [40]. Zuccon and Koopman [41], and later Zuccon [42], have proposed and investigated a family of measures based on the gain-discount framework, where the gain of a document is influenced by both its topical relevance and its understandability. They showed that although generally correlated, topical relevance evaluation alone provides differing system rankings compared with understandability-biased evaluation measures. In this study, we further explored the development of retrieval methods that combine signals about topical relevance and understandability.

Methods

Data Collection

In this paper, we investigated methods to estimate Web page understandability, including the effect that HTML preprocessing pipelines and heuristics have, and their search effectiveness when employed within retrieval methods. To obtain both topical relevance and understandability assessments, we used the data from the CLEF 2015 and 2016 eHealth collections. The CLEF eHealth initiative is a research community–shared task aimed at creating resources for evaluating health search engines aimed at the general public [43]. Note, in the remainder of this paper, we refer to topical relevance simply as relevance, when this does not cause confusion.

The CLEF 2015 collection contains 50 queries and 1437 documents that have been assessed as relevant by clinical experts and have an assessment for understandability [44]. Documents in this collection are a selected crawl of health websites, of which the majority are certified HON websites. The CLEF 2016 collection contains 300 queries and 3298 relevant documents that also have been assessed with respect to understandability [14]. Documents in this collection belong to the ClueWeb12 B13 corpus [45], and thus are general English Web pages, not necessarily targeted to health topics nor of a controlled quality (as are the HON certified pages). Understandability assessments were provided on a 5-point Likert scale for CLEF 2015 on a 0 to 100 range for CLEF 2016 (0 indicates the highest understandability).

To support the investigation of methods to automatically estimate the understandability of Web pages, we further considered correlations between multiple human assessors (interassessor agreement). For CLEF 2015, we used the publicly available additional assessments made by unpaid medical students and health consumers collected by Palotti et al [46] in a study of how medical expertise affects assessments. For CLEF 2016, we collected understandability assessments for 100 documents. In total, 3 members of our research team, who did not author this paper and are not medical experts, were recruited to provide the assessments (the correlation of these additional assessments and CLEF's ground truth is examined further in this paper). The Relevation tool [47] was used to assist with the assessments, mimicking the settings used in CLEF.

Understandability Estimators

Several methods have been used to estimate the understandability of health Web pages, with the most popular methods (at least in the biomedical literature) being RF based on surface level characteristics of the text. Next, we outline the categories of methods to estimate understandability used in this study; an overview is shown in Textboxes 1 to 10.

Traditional Readability Formulae

These include the most popular RF [25-27] as well as other less popular ones [48-51]. An extensive description of these RF is provided in surveys by Collins-Thompson [52] and Dubay [28]. A complete list of methods is provided in Textbox 1.

Raw Components of Readability Formulae

These are formed by the *building blocks* used in the traditional RF. Examples include the average number of characters per word and the average number of syllables in a sentence. Words are divided into syllables using the Python package Pyphen [53]. A complete list of methods is provided in Textbox 2.

General Medical Vocabularies

These include methods that count the number of words with a medical prefix or suffix, that is, beginning or ending with Latin or Greek particles (eg, amni-, angi-, algia-, and arteri-), and text strings included in lists of acronyms or in medical vocabularies such as the International Statistical Classification of Diseases and Related Health Problems (ICD), Drugbank and the OpenMedSpel dictionary [54]. An acronym list from the ADAM database [55] was used. Methods in this category were matched with documents using simple keyword matching. A complete list of methods is provided in Textbox 3.

Consumer Medical Vocabulary

The popular MetaMap [56] tool was used to map the text content of Web pages to entries in CHV [34]. We used the MetaMap semantic types to retain only concepts identified as symptoms or diseases. Similar approaches have been commonly used in the literature [57-60]. A complete list of methods is provided in Textbox 4.

Expert Medical Vocabulary

Similar to the CHV features, we used MetaMap to convert the content of Web pages into MeSH entities, studying symptom and disease concepts separately. A complete list of methods is provided in Textbox 5.

Natural Language Features

These included commonly used natural language heuristics such as the ratio of POS classes, the height of the POS parser tree, the number of entities in the text, the sentiment polarity [61], and the ratio of words found in English vocabularies. The Python package Natural Language Toolkit [62] was used for sentiment analysis, POS tagging, and entity recognition. The GNU Aspell [63] dictionary was used as a standard English vocabulary and a stop word list was built by merging those of Indri [64] and Terrier [65]. Discourse features, such as the distribution of POS classes and density of entity in a text, were previously studied in the task of understandability prediction [66] and found superior to complex features such as entity coreference and entity grid [67]. To the best of our knowledge, sentiment polarity was never investigated in this task. Our intuition is that the content produced by laypeople in patient forums or blogs (easy to read) is potentially more emotional than scientific publications (hard to read). A complete list of methods is provided in Textbox 6.

HTML Features

These include the identification of a large number of HTML tags, which were extracted with the Python library BeautifulSoup [68]. The intuition for these features is that Web pages with many images and tables might explain and summarize health content better, thus providing more understandable content to the general public. A complete list of methods is provided in Textbox 7.

Textbox 1. Readability formulae (RF) used to estimate understandability.

Readability feature

- Automated Readability Index [48]
- Coleman-Liau Index (CLI) [25]
- Dale-Chall Index (DCI) [26]
- Flesch-Kincaid Grade Level [27]
- Flesch Reading Ease (FRE) [27]
- Gunning Fog Index (GFI) [49]
- Lasbarhetsindex (LIX) [50]
- Simple Measure of Gobbledygook (SMOG) [51]

Textbox 2. Raw components of readability formulae (CRF) used to estimate understandability. For all features, raw values, values normalized by number of words in a document, and values normalized by number of sentences in a document were used.

Components of readability feature

- # of Characters
- # of Words
- # of Sentences
- # of Difficult Words (Dale-Chall list [26])
- # of Words Longer than 4 Characters
- # of Words Longer than 6 Characters
- # of Words Longer than 10 Characters
- # of Words Longer than 13 Characters
- # of Number of Syllables
- # of Polysyllable Words (>3 Syllables)

Textbox 3. General medical vocabulary features used to estimate understandability. For all features, raw values, values normalized by number of words in a document, and values normalized by number of sentences in a document were used.

General medical vocabularies (GMVs)

- # of words with medical prefix
- # of words with medical suffix
- # of acronyms
- # of International Statistical Classification of Diseases and Related Health Problems (ICD) concepts
- # of Drugbank
- # of words in medical dictionary (OpenMedSpel)

Textbox 4. Consumer medical vocabulary features used to estimate understandability. For all features, raw values, values normalized by number of words in a document, and values normalized by number of sentences in a document were used.

Consumer medical vocabularies (CMV)

- Consumer health vocabulary (CHV) mean score for all concepts
- # of CHV concepts
- CHV mean score for symptom concepts
- # of CHV symptom concepts
- CHV mean score for disease concepts
- # of CHV disease concepts

Textbox 5. Expert medical vocabulary features used to estimate understandability. For all features, raw values, values normalized by number of words in a document, and values normalized by number of sentences in a document were used.

Expert medical vocabulary (EMV)

- # of Medical Subject Headings (MeSH) concepts
- Average tree of MeSH concepts
- # of MeSH symptom concepts
- Average tree of MeSH symptom concepts
- # of MeSH disease concepts
- Average tree of MeSH disease concepts

Textbox 6. Natural language features used to estimate understandability. For all features, raw values, values normalized by number of words in a document, and values normalized by number of sentences in a document were used.

Natural language features (NLF)

- Positive words
- Negative words
- Neutral words
- # of verbs
- # of nouns
- # of pronouns
- # of adjectives
- # of adverbs
- # of adpositions
- # of conjunctions
- # of determiners
- # of cardinal numbers
- # of particles or other function words
- # of other part of speech (POS; foreign words and typos)
- # of punctuation
- # of entities
- Height of POS parser tree
- # of stop words
- # of words not found in Aspell Engish dictionary
- Average tree of Medical Subject Headings (MeSH) disease concepts

Word Frequency Features

Generally speaking, common and known words are usually frequent words, whereas unknown and obscure words are generally rare. This idea is implemented in RF such as the DCI, which uses a list of common words and counts the number of words that fall outside this list (complex words) [26] and has shown success in other recent approaches [69,70]. We extended these observations by studying corpus-wide word frequencies. In total, 3 corpora were analyzed to extract word frequencies:

- Medical Reddit: Reddit [71] is a Web forum with a sizeable user community, which is responsible for generating and moderating its content. This forum is intensively used for health purposes, for example, in the Reddit community AskDocs [72], licensed nurses and doctors (subject to user identity verification) advise help seekers free of charge. We selected 6 of such communities (medical, AskDocs, AskDoctorSmeeee, Health, WomensHealth, and Mens_Health) and downloaded all user interactions available until September 1, 2017, using the Python library Python Reddit Wrapper PRAW [73]. In total, 43,019 discussions were collected.
- Medical English Wikipedia: after obtaining a recent Wikipedia dump [74] (May 1, 2017), we filtered papers to only those containing an Infobox in which at least one of the following words appeared as a property: ICD10, ICD9, DiseasesDB, MeSH, MeSHID, MeshName, MeshNumber, GeneReviewsName, Orphanet, eMedicine, MedlinePlus, drug_name, Drugs.com, DailyMedID, and LOINC. A Wikipedia infobox is a structured template that appears on the right of Wikipedia pages summarizing key aspects of papers. This process followed the method by Soldaini et al [75], which favors precision over recall when identifying a health-related paper. This resulted in a collection of 11,868 papers.
- PubMed Central: PubMed Central is a Web-based database of biomedical literature. We used the collection distributed for the Text Retrieval Conference (TREC) 2014 and 2015 Clinical Decision Support Track [76,77], consisting of 733,191 papers.

Textbox 7. HTML features used to estimate understandability.

HTML features (HF)

- # of abbreviation (abbr tags)
- # of links (A tags)
- # of blockquote tags
- # of bold tags
- # of cite tags
- # of divisions or sections (div tags)
- # of forms tags
- # of heading H1 tags
- # of heading H2 tags
- # of heading H3 tags
- # of heading H4 tags
- # of heading H5 tags
- # of heading H6 tags
- Total # of headings (any heading H above)
- # of image tags
- # of input tags
- # of link tags
- # of description lists (DL tags)
- # of unordered lists (UL tags)
- # of ordered lists (OL tags)
- Total # of any list (DL+UL+OL)
- # of short quotations (Q tags)
- # of scripts tags
- # of spans tags
- # of table tags
- # of paragraphs (P tags)

A summary of the statistics of the corpora is reported in Table 1. We modeled word frequencies in a corpus in a straightforward manner: we sorted the word frequencies and normalized word rankings such that values close to 100 are attributed to common words and values close to 0 to rare words. Thereafter, we replaced each word in a document by a number ranging from 0 to 100, which represents the frequency of that word in the corpus. Finally, we extracted features based on the word frequency distribution for that document. For example, the feature *75th percentile English Wikipedia* is a number between 0 and 100 representing how frequent is the word at the 75th percentile of a document in which word frequencies were extracted from the English Wikipedia corpus. Unless explicitly stated otherwise, we ignored out-of-vocabulary (OV) words in the corpus. A complete list of methods is provided in Textbox 8.

Machine Learning on Text-Regressors and Classifiers

These include machine learning methods for estimating Web page understandability. Although Collins-Thompson highlighted the promise of estimating understandability using machine learning methods, a challenge is identifying the background corpus to be used for training [52]. To this aim, we used the 3 corpora detailed above, and assumed understandability labels according to the expected difficulty of documents in these collections:

- Medical Reddit (label 1): Documents in this corpus are expected to be written in a colloquial style, and thus the easiest to understand. All the conversations are, in fact, explicitly directed to assist inexpert health consumers
- Medical English Wikipedia (label 2): Documents in this corpus are expected to be less formal than scientific papers, but more formal than a Web forum like Reddit, thus somewhat more difficult to understand
- PubMed Central (label 3): Documents in this corpus are expected to be written in a highly formal style, as the target audience are physicians and biomedical researchers.

Table 1. Statistics for the corpora used as background models for understandability estimations.

Statistics	Medical Wikipedia	Medical Reddit	PubMed Central
Documents, n	11,868	43,019	733,191
Words, n	10,655,572	11,978,447	144,024,976
Unique words, n	467,650	317,106	2,933,167
Average words per document, mean (SD)	898.90 (1351.76)	278.45 (359.70)	227.22 (270.44)
Average characters per document, mean (SD)	5107.81(7618.57)	1258.44 (1659.96)	1309.11(1447.31)
Average characters per word, mean (SD)	5.68 (3.75)	4.52 (3.52)	5.76 (3.51)

Textbox 8. Word frequency features used to estimate understandability.

Word frequency features (WFF)

- 25th percentile English Wikipedia
- 50th percentile English Wikipedia
- 75th percentile English Wikipedia
- Mean rank English Wikipedia
- Mean rank English Wikipedia—includes out-of-vocabulary (OV) words
- 25th percentile Medical Reddit
- 50th percentile Medical Reddit
- 75th percentile Medical Reddit
- Mean rank Medical Reddit
- Mean rank Medical Reddit—includes OV
- 25th percentile Pubmed
- 50th percentile Pubmed
- 75th percentile Pubmed
- Mean rank Pubmed
- Mean rank Pubmed—includes OV
- 25th percentile Wikipedia+Reddit+Pubmed
- 50th percentile Wikipedia+Reddit+Pubmed
- 75th percentile Wikipedia+Reddit+Pubmed
- Mean rank Wikipedia+Reddit+Pubmed
- Mean rank Wikipedia+Reddit+Pubmed—includes OV

Textbox 9. Machine learning regressor features used to estimate understandability.

Machine learning regressors (MLR)

- Linear regressor
- Multilayer perceptron regressor
- Random forest regressor
- Support vector machine regressor
- eXtreme Gradient Boosting Regressor

Textbox 10. Machine learning classifier features used to estimate understandability.

Machine learning classifiers (MLC)

- Logistic regression
- Multilayer perceptron classifier
- Random forest classifier
- Support vector machine classifier
- Multinomial naive Bayes
- eXtreme Gradient Boosting Classifier

On the basis of the labels of each class above, models were learnt using all documents from these corpora after features were extracted using latent semantic analysis with ten dimensions on top of TF-IDF calculated for each word. We modeled a classification task as well as a regression task using these corpora. In the classification task, the first step is to train a classifier on documents belonging to these three collections with the three different classes shown above. The second step is to use the classifier to estimate which of these three possible classes an unseen document from the CLEF 2015 or CLEF 2016 would belong. Similarly, in the regression task, after training, a regressor has to estimate an understandability value to an unseen CLEF document. We hypothesize that documents that are more difficult to read are more similar to PubMed documents than to Wikipedia or Reddit ones. A complete list of methods is provided in Textboxes 9 and 10.

Preprocessing Pipelines and Heuristics

As part of our study, we investigated the influence that the preprocessing of Web pages had on the estimation of understandability computed using the methods described above. We did so by comparing the combination of a number of preprocessing pipelines, heuristics, and understandability estimation methods with human assessments of Web page understandability. Our experiments extended our previous work [20] and provided a much more thorough analysis, as they only evaluated surface level RF and did not compare their results against human assessments.

To extract the content of a Web page from the HTML source we tested: BeautifulSoup, *Naive* [68], which just naively removes HTML tags and Boilerpipe, *Boi* [78] and Justext, *Jst* [79], which eliminates boilerplate text together with HTML tags. Our data analysis in Palotti et al [20] highlighted that the text in HTML fields such as titles, menus, tables, and lists often missed a correct punctuation mark, and thus, the text extracted from them could be interpreted as many short sentences or few very long sentences, depending on whether a period was forced at the end of fields or sentences. We, thus, implemented the same 2 heuristics devised to deal with this: *ForcePeriod (FP)* and *DoNotForcePeriod (DNFP)*. If a punctuation mark is found at the end of a field or sentence, it is kept as it is. However, if no punctuation mark is found at the end of a field or sentence, the FP heuristic forces the insertion of a period at the end of that extracted HTML field, whereas the DNFP does not.

Integrating Understandability into Retrieval

We then investigated how understandability estimations can be integrated into retrieval methods to increase the quality of search results. Specifically, we considered 3 retrieval methods of differing quality for the initial retrieval. These included the best 2 runs submitted to each CLEF task, and a plain BM25 baseline (default Terrier parameters: b=0.75 and k_1=1.2). BM25 is a probabilistic term weighting scheme commonly used in information retrieval and is defined with respect to the frequency of a term in a document, the collection frequency of that term, and the ratio between the length of the document and the average document length. As understandability estimators, we used the XGB regressor [80] as well as Simple Measure of Gobbledygook (SMOG) for CLEF 2015 and DCI for CLEF 2016. These were selected as they were the best performing RF and machine learning methods for each collection (details on the evaluation of understandability estimators presented in the Results section). Remember that, as described in the *Related Work* section, RF are a specific approach to estimate understandability. Note that in XGB, for assessed documents we used 10-fold cross validation, training XGB on 90% of the data, and used its predictions for the remaining 10%. For unassessed documents, we trained XGB on all assessed data and applied this model to generate predictions. Different machine learning methods and feature selection schemes were experimented with; results are available in the Multimedia Appendix 1. XGB was selected because its results were the best among the machine learning test (which include all machine learning methods listed in Textboxes 9 and 10).

To integrate understandability estimators into the retrieval process, we first investigated *reranking* search results retrieved by the initial runs purely based on the understandability estimations. If all the search results from a run were to be considered, then such a reranking method might place at early ranks Web pages highly likely to be understandable, but possibly less likely to be topically relevant. To balance relevance and understandability, we only reranked the first *k* documents. We explored rank cut-offs k=15,20,50. As evaluation was performed with respect to the first n=10 rank positions, the setting k=15 provided a conservative reranking of search results, whereas, k=50 provided a less conservative reranking approach.

Table 2. Learning to rank settings.

Strategy	Explanation	Labeling function	
		CLEF[a] 2015	CLEF 2016
LTR[b] 1	Model built *only* on the topicality labels with IR[c] features	F[d](R[e],U[f])=R	F(R,U)=R
LTR 2	Model built *only* on the topicality labels with IR and understandability features	F(R,U)=R	F(R,U)=R
LTR 3	Model combines understandability and topicality labels. Uses IR and understandability features	F(R,U)=R×U/3	F(R,U)=R×(100-U)/100
LTR 4	Model built *only* on easy-to-read documents. Uses IR and understandability Features	F(R,U)=R, if U≥2 F(R,U)=0, otherwise	F(R,U)=R, if U≤40 F(R,U)=0, otherwise
LTR 5	Model built boosting easy-to-read documents. Uses IR and understandability Features	F(R,U)=2×R, if U≥2 F(R,U)=R, otherwise	F(R,U)=2×R, if U≤40 F(R,U)=R, otherwise

[a]CLEF: Conference and Labs of the Evaluation Forum.

[b]LTR: learning to rank.

[c]IR: information retrieval.

[d]F: function.

[e]R: topical relevance of a document.

[f]U: understandability.

As an alternative to the previous 2-step ranking strategy for combining topical relevance and understandability, we explored the *fusion* of 2 search result lists separately obtained for relevance and understandability. For this, we used the reciprocal rank fusion method [81], which was shown effective for combining 2 lists of search results based on their documents' *ranks*, rather than scores. This approach was selected above score-based fusion methods because the distribution of relevance scores for the retrieved documents differed sensibly (both in magnitude and spread) with that of understandability scores: in such a case, score-based fusion is not appropriate. For relevance, we used, separately, the 3 retrieval methods for each collection. For CLEF 2015, we used BM25 and the submissions made by the East China Normal University (ECNU) team [82] and the Korean Institute of Science and Technology Information (KISTI) team [83]. For CLEF2016, we also used BM25 and the submissions made by the Georgetown University Information Retrieval (GUIR) team [84] and ECNU [85]. For understandability, we used, separately, the estimations from SMOG or DCI and XGB. Moreover, for this approach, we studied limiting the ranking of results to be considered by the methods across the cut-offs k=15,20,50.

Finally, we considered a third alternative to combine relevance and understandability: using *learning to rank* with features derived from retrieval methods (information retrieval (IR) features) and understandability estimators. Learning to rank refers to a family of machine learning methods where ranking models are learnt from training data (and associated features). With the CLEF 2015 and 2016 collections, we explored 5 combinations of label attribution and feature sets, maintaining the same pairwise learning to rank algorithm based on tree boosting (XGB). These combinations are listed in Table 2, with R being the relevance of documents and U their understandability estimation. Although the definitions of learning to rank (LTR) 1 and LTR 2 are straightforward, the other methods deserve some further explanation. In LTR 3, a penalty was proportionally assigned to documents according to their understandability score U. For example, for CLEF 2016, a document with understandability U=0 received no penalty, as 0 was the easiest level of understanding, whereas another with understandability 50 received a 50% penalty, meaning that its relevance score was halved. LTRs 4 and 5 were based on a fixed threshold applied to the understandability score: if the score was higher than the threshold (U=2 for CLEF 2015 and U=40 for CLEF 2016), then the original relevance score (for LTR 4) or a boosted value (for LTR 5) was assigned to the corresponding document. We used the thresholds U=2 for CLEF 2015 and U=40 for CLEF 2016, based on the distribution of understandability assessments and the semantic of understandability labels [44,14].

Evaluation Measures

In the experiments, we used Pearson, Kendall, and Spearman correlations to compare the understandability assessments of human assessors with estimations obtained by the considered automated approaches, under all combinations of pipelines and heuristics. Pearson correlation is used to calculate the strength of the linear relationship between 2 variables, whereas Kendall and Spearman measure the rank correlations among the variables. We opted to report all 3 correlation coefficients to allow for a thorough comparison with other work, as they are equally used in the literature.

For the retrieval experiments, we used evaluation measures that rely on both (topical) relevance and understandability. The uRBP measure [42] extends rank biased precision (RBP) to situations where multiple relevance dimensions are used. The measure is formulated as $uRBP(p)=(1-p) \times \sum^{k} p^{k-1} \times r(d@k) \times u(d@k)$, where r(d@k) is the gain for retrieving a relevant document at rank k and u(d@k) is the gain for retrieving a document of a certain understandability at rank k; p is the RBP persistence parameter. This measure was an official evaluation measure used in CLEF (we also set *P*=.8).

A drawback of uRBP is that relevance and understandability are combined into a unique evaluation score, thus making it difficult to interpret whether improvements are because of more understandable or more topical documents being retrieved. To overcome this, we used the multidimensional metric (MM) framework introduced by Palotti et al [86] which first separately calculates an RBP value for relevance and another for understandability, and then combine them into a unique effectiveness measure:

- $RBP_r@n(p)$: uses the relevance assessments for the top n search results (ie, this is the common RBP). We regarded a document as topically relevant if assessed as somewhat relevant or highly relevant.
- $RBP_u@n(p)$: uses the understandability assessments for the top n search results. We regarded a document as understandable (1) for CLEF 2015 if assessed easy or somewhat easy to understand and (2) for CLEF 2016 if its assessed understandability score was smaller than a threshold U. We used U=40, based on the distribution of understandability assessments. Assessors were presented with a slider for understandability assessment and U=50 was labeled as average understandability. This created a bimodal distribution of understandability assessments, with U=40 being a good upper limit for easy-to-read documents. The understandability distribution can be found in the Multimedia Appendix 2.
- $MM_{RBP}@n(p)=2\times(RBP_r@n\times RBP_u@n)/(RBP_r@n+RBP_u@n)$: combines the previous 2 RBP values into a unique measurement using the harmonic mean (in the same fashion that the F_1 measure combines recall and precision).

For all measures, we set n=10 because shallow pools were used in CLEF along with measures that focused on the top 10 search results (including $RBP_r@10$). Shallow pools refer to the selection of a limited number of documents to be assessed for relevance, among the documents retrieved at the top ranks by a search engine.

Along with these measures of search effectiveness, we also recorded the number of unassessed documents, the RBP residuals, $RBP^*_r@10$, $RBP^*_u@10$, and MM^*_{RBP}, that is, the corresponding measures calculated by ignoring unassessed documents. These latter measures implement the condensed measures approach proposed by Sakai as a way to deal with unassessed documents [87]. We did this to minimize pool bias as the pools built in CLEF were of limited size and the investigated methods retrieved a substantial number of unassessed documents. Pool bias refers to the possible bias in the evaluation toward systems that have contributed documents to the assessment pool: these erroneously receive higher evaluation scores compared with systems that did not contribute to the pool (ie, that were not sampled to create the set of documents to be judged for relevance).

Results

Evaluation of Understandability Estimators

To keep this paper succinct, in the following we only report a subset of the results. The remaining results (which show similar trends to those reported here) are made available in the Multimedia Appendix 3 material for completeness.

Using the CLEF eHealth 2015 and 2016 collections, we studied the correlations of methods to estimate Web page understandability compared with human assessments. For each category of understandability estimation, Tables 3 and 4 report the methods with highest Pearson and Spearman or Kendall correlations for CLEF 2015 and 2016, respectively. For each method, we used the best preprocessing settings; a study of the impact of preprocessing is reported in the next subsection.

Overall, Spearman and Kendall correlations obtained similar results (in terms of which methods exhibited the highest correlations): this was expected as, unlike Pearson, they are both rank-based correlations.

For traditional RF, SMOG had the highest correlations for CLEF 2015 and DCI for CLEF 2016, regardless of correlation measure. These results resonate with those obtained for the category of raw components of readability formulae (CRF). In fact, the polysyllable words measure, which is the main feature used in SMOG, had the highest correlation for CLEF 2015 among methods in this category. Similarly, the number of difficult words, which is the main feature used in DCI, had the highest correlation for CLEF 2016 among methods in this category.

When examining the expert vocabulary category (EMV), we found that the number of MeSH concepts obtained the highest correlations with human assessments; however, its correlations were substantially lower than those achieved by the best method from the consumer medical vocabularies category (CMV), that is, the scores of CHV concepts. For the natural language category (NLF), we found that the number of pronouns, the number of stop words, and the number of OV words had the highest correlations—and these were even higher than those obtained with MeSH- and CHV-based methods. In turn, the methods that obtained the highest correlations among the HTML category (HF) and counts of P tags and list tags exhibited overall the lowest correlations compared with methods in the other categories. P tags are used to create paragraphs in a Web page, being thus a rough proxy for text length. Among methods in the word frequency category (WFF), the use of Medical Reddit (but also of PubMed) showed the highest correlations, and these were comparable with those obtained by the RF.

Finally, regressors (MLR) and classifiers (MLC) exhibited the highest correlations among all methods: in this category, the XGB regressor and the multinomial Naive Bayes best correlated with human assessments.

Table 3. Methods with the highest correlation per category for Conference and Labs of the Evaluation Forum (CLEF) 2015.

Category	Method	Preprocessing	Pearson	Spearman	Kendall
Readability formulae	Simple Measure of Gobbledygook Index	Jst Do Not Force Period (DNFP)	*.438* [a]	*.388*	*.286*
Components of readability formulae (CRF)	Average number of Polysyllables words per sentence	Jst force period (FP)	*.429*	.364	.268
CRF	Average number of Polysyllables words per sentence	Jst DNFP	.192	*.388*	*.286*
General medical vocabularies (GMVs)	Average number of medical prefixes per word	Naïve FP	*.314*	.312	.229
GMVs	Number of medical prefixes	Naïve FP	.131	*.368*	*.272*
Consumer medical vocabulary (CMV)	Consumer health vocabulary (CHV) mean score for all concepts	Naïve FP	*.371*	*.314*	*.228*
Expert medical vocabulary (EMV)	Number of medical concepts	Naïve FP	*.227*	*.249*	*.178*
Natural language features (NLF)	Number of words not found in Aspell dictionary	Jst DNFP	*.351*	.276	.203
NLF	Number of pronouns per word	Naïve FP	.271	*.441*	*.325*
HTML features (HF)	Number of P tags	None	*.219*	*.196*	*.142*
Word frequency features (WFF)	Mean rank Medical Reddit	Jst DNFP	*.435*	.277	.197
WFF	25th percentile Pubmed	Jst DNFP	.330	*.347*	*.256*
Machine learning regressors (MLR)	eXtreme Gradient Boosting (XGB) Regressor	Boi DNFP	*.602*	.394	.287
MLR	XGB Regressor	Jst FP	.565	*.438*	*.324*
Machine learning classifiers	Multinomial Naïve Bayes	Naïve FP	*.573*	*.477*	*.416*

[a]Italics used to highlight the best result of each group.

Table 4. Methods with the highest correlation per category for Conference and Labs of the Evaluation Forum (CLEF) 2016.

Category	Method	Preprocessing	Pearson	Spearman	Kendall
Readability formulae (RF)	Dale-Chall Index (DCI)	Jst force period (FP)	*.439* [a]	.381	*.264*
RF	DCI	Boi FP	.437	*.382*	*.264*
Components of readability formulae (CRF)	Average number of difficult word per Word	Boi FP	*.431*	*.379*	*.262*
General medical vocabularies (GMVs)	Average prefixes per sentence	Jst FP	*.263*	.242	.164
GMVs	International Statistical Classification of Diseases and Related Health Problems concepts per sentence	Jst do not force period (DNFP)	.014	*.253*	*.172*
Consumer medical vocabulary (CMV)	Consumer health vocabulary (CHV) mean score for all concepts	Jst FP	*.329*	.313	.216
CMV	CHV mean score for all concepts	Boi FP	*.329*	*.325*	*.224*
EMV	Number of MeSH (Medical Subject Headings) concepts	Boi DNFP	*.201*	.166	.113
Expert medical vocabulary (EMV)	Number of MeSH disease concepts	Boi DNFP	.179	*.192*	*.132*
Natural language features (NLF)	Average stop word per word	Boi FP	*.344*	.312	.213
NLF	Number of pronouns	Boi FP	.341	*.364*	*.252*
HTML features (HF)	Number of lists	None	*.114*	.021	.015
HF	Number of P tags	None	.110	*.123*	*.084*
Word frequency features (WFF)	Mean rank Medical Reddit	Boi DNFP	*.387*	.312	.214
WFF	50th percentile Medical Reddit	Jst DNFP	.351	*.315*	*.216*
Machine learning regressors (MLR)	eXtreme Gradient Boosting (XGB) Regressor	Jst DNFP	*.454*	*.373*	.258
MLR	Random Forest Regressor	Boi DNFP	.389	.355	*.264*
Machine learning classifiers	Multinomial Naïve Bayes	Jst FP	*.461*	*.391*	*.318*

[a]Italics used to highlight the best result of each group.

Evaluation of Preprocessing Pipelines and Heuristics

Results from experiments with different preprocessing pipelines and heuristics are shown in Figures 2 and 3, respectively for CLEF 2015 and 2016. For each category of methods and combination of preprocessing and heuristics, we report their variability in terms of Spearman rank correlation with human assessments. Results for Pearson and Kendall correlations are reported in the Multimedia Appendix 3, but showed similar trends. We further report the summary results across all understandability assessment methods and sentence-ending heuristics for each of the preprocessing pipelines. Finally, we also report the interassessor correlation (last box) when multiple assessors provided judgments about the understandability of Web pages. This provides an indication of the range of variability and subjectiveness when assessing understandability, along with the highest correlation we measured between human assessors.

We first examined the correlations between human assessments and RF. We found that the *Naive* preprocessing resulted in the lowest correlations, regardless of RF and heuristic (although *DoNotForcePeriod* performed better than *ForcePeriod*). Using Justext or Boilerplate resulted in higher correlations with human understandability assessments, and the *ForcePeriod* heuristic was shown to be better than *DoNotForcePeriod*. These results confirm our hypotheses in Palotti et al [20]: we found these settings to produce lower variances in understandability estimations, and thus hypothesized that they were better suited to the task.

Overall, among RF, the best results (highest correlations) were obtained by SMOG and DCI (see also Tables 3 and 4). Although no single setting outperformed the others in both collections, we found that the use of CLI and FRE with *Justext* provided the most stable results across the collections, with correlations as high as the best ones in both collections. These results confirmed our previous advice [20], that is, in general, if using readability measures, CLI is to be preferred, along with an appropriate HTML extraction pipeline, regardless of the heuristic for sentence ending. We provide detailed plots to compare the results in this paper with those in Palotti et al [20] in the Multimedia Appendix 4.

When considering methods beyond those based on RF, we found that the highest correlations were achieved by the regressors (MLR) and classifiers (MLC), independently of the preprocessing method used. There is little difference in terms of effectiveness of methods in these categories, with the exception of regressors on CLEF 2015 that exhibited not negligible variances: whereas for the neural network regressor the Pearson correlation was .44 and for the support vector regressor it was only .30.

Figure 2. Correlations between understandability estimators and human assessments for Conference and Labs of the Evaluation Forum 2015. For example, the first boxplot on the top represents the distribution of Spearman correlations with human assessments across all features in the category readability formulae, obtained with the Naive Force Period preprocessing. Each box extends from the lower to the upper quartile values, with the red marker representing the median value for that category. Whiskers show the range of the data in each category and circles represent values considered outliers for the category (eg, Spearman correlation for Simple Measure of Gobbledygook (SMOG) index was .296 and for Automated Readability Index (ARI) was .194: these were outliers for that category). CMV: consumer medical vocabulary; CRF: components of readability formulae; DNFP: Do Not Force Period; EMV: expert medical vocabulary; FP: Force Period; GMV: general medical vocabulary; MLC: machine learning classifiers; MLR: machine learning regressors; NLF: natural language features; RF: readability formulae; WFF: word frequency features.

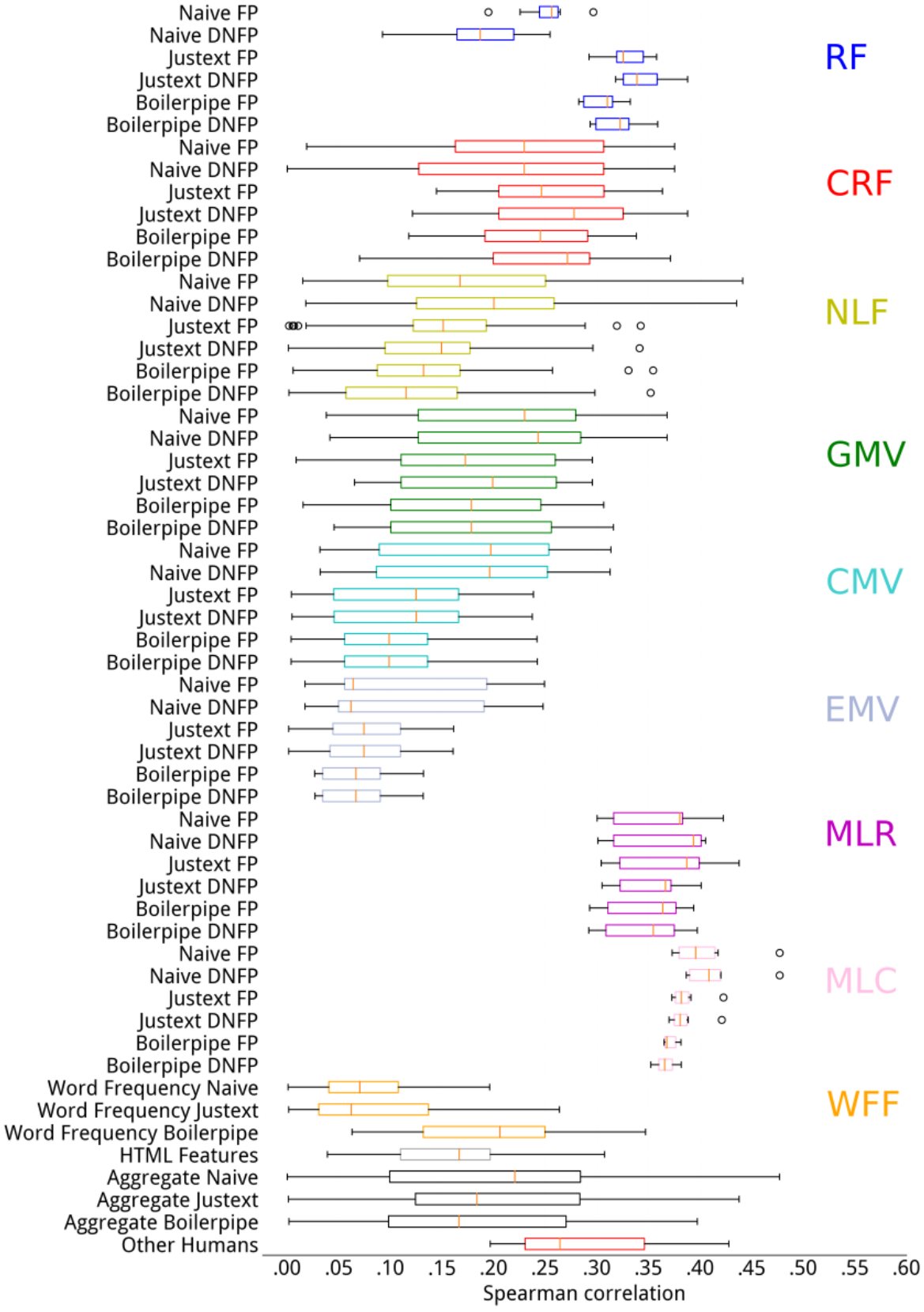
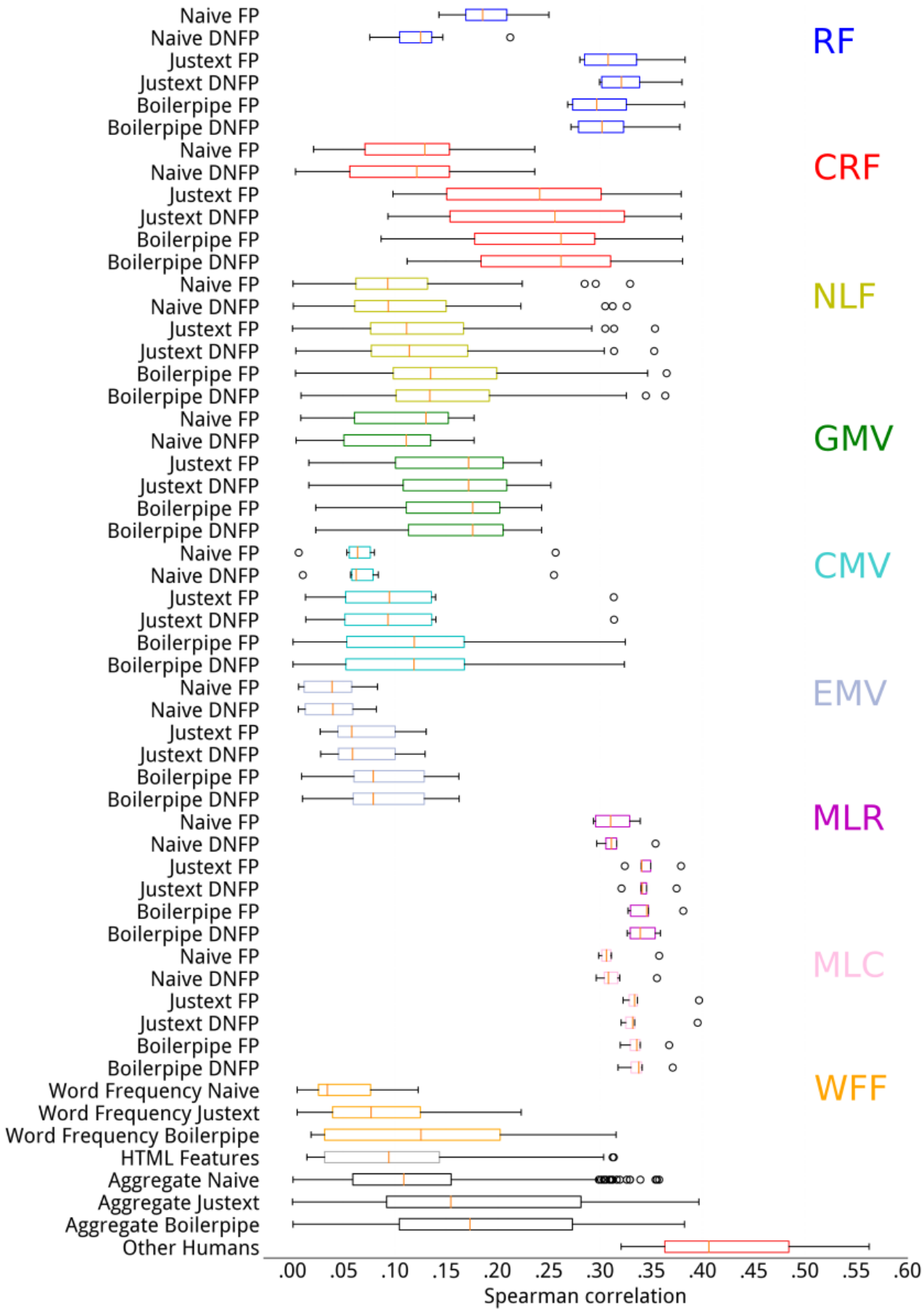


Figure 3. Correlations between understandability estimators and human assessments for Conference and Labs of the Evaluation Forum (CLEF) 2016. CMV: consumer medical vocabulary; CRF: components of readability formulae; DNFP: Do Not Force Period; EMV: expert medical vocabulary; FP: Force Period; GMV: general medical vocabulary; MLC: machine learning classifiers; MLR: machine learning regressors; NLF: natural language features; RF: readability formulae; WFF: word frequency features.



A common trend when comparing preprocessing pipelines is that the Naïve pipeline provided the weakest correlations with human assessments for CLEF 2016, regardless of estimation methods and heuristics. This result, however, was not confirmed for CLEF 2015, where the Naive preprocessing negatively influenced correlations for the RF category, but not for other categories, although it was generally associated with larger variances for the correlation coefficients.

Evaluation of Understandability Retrieval

Reranking Experiments

Results for the considered retrieval methods are reported in Figures 4-8. We report only the results for CLEF 2016 for brevity; those for CLEF 2015 exhibited similar trends and are included in the Multimedia Appendix 5. When reranking results, we risk bringing to the top position a document that was never assessed. The RBP residuals (shown in gray in Figures 3-8) show the possible gains that unassessed documents can have on the evaluation, as it assumed that all unassessed documents are relevant. Another way to quantify the effect that unassessed documents have on evaluation is looking at the average number of unassessed documents in the top 10 results: this is given by the metric *Unj@10*. Larger values of *Unj@10* imply that actual effectiveness might be noticeably larger. Here, we also show the values for the condensed measures.

The effectiveness of the top 2 submissions to CLEF 2016 and the BM25 baseline are reported in Figure 4. In turn, we report the results of each subexperiment: *simple reranking* (Figures 5 and 6), *fusion experiments* (Figure 7), and *learning to rank* (Figure 8).

Figure 4. Baseline results for the best 2 submissions to Conference and Labs of the Evaluation Forum (CLEF) 2016 (Georgetown University Information Retrieval [GUIR] and East China Normal University [ECNU]) and the Best Match 25 (BM25) baseline of Terrier. MM: multidimensional metric; RBP: rank biased precision.

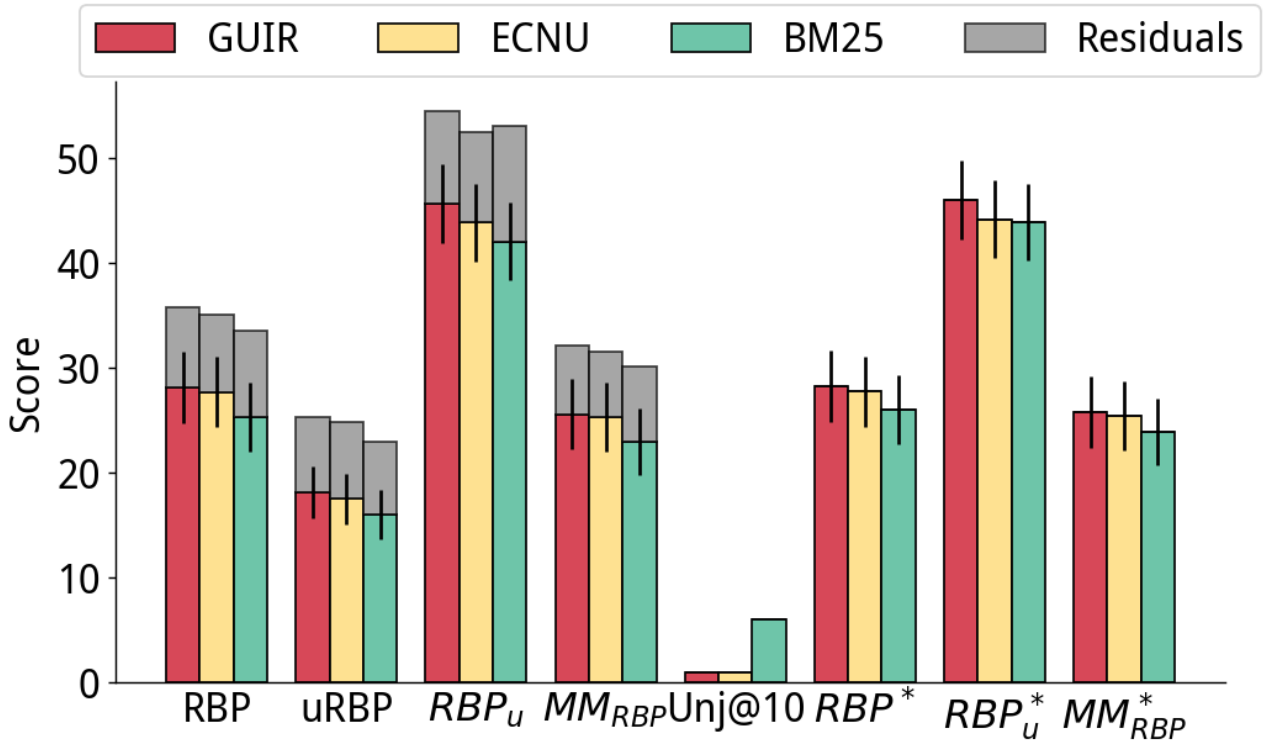


Figure 5. Reranking of the runs based on the Dale-Chall readability formula. ECNU: East China Normal University; GUIR: Georgetown University Information Retrieval; MM: multidimensional metric; RBP: rank biased precision.

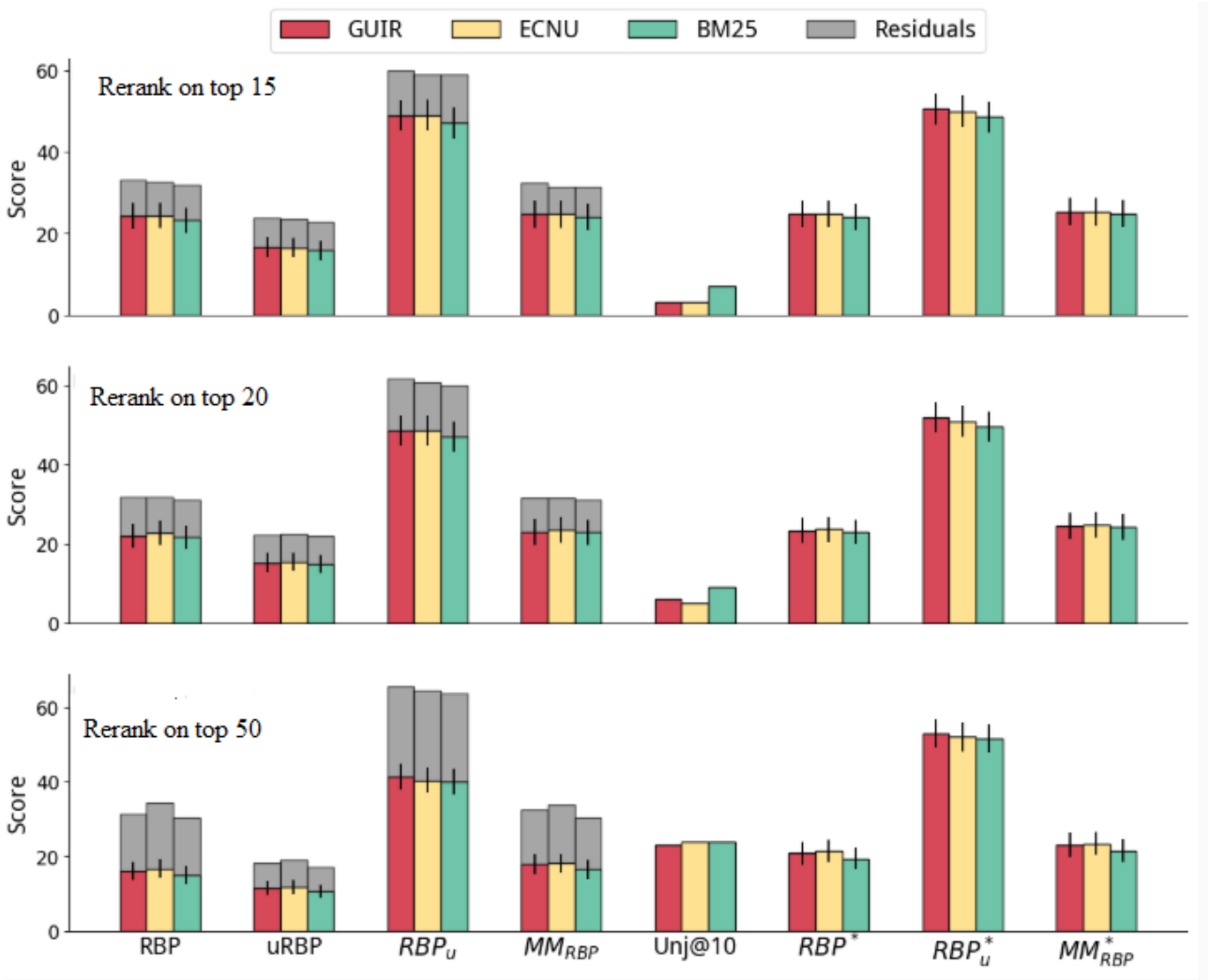


Figure 6. Reranking of the runs based on the eXtreme Gradient Boosting (XGB) regressor to estimate understandability. ECNU: East China Normal University; GUIR: Georgetown University Information Retrieval; MM: multidimensional metric; RBP: rank biased precision.

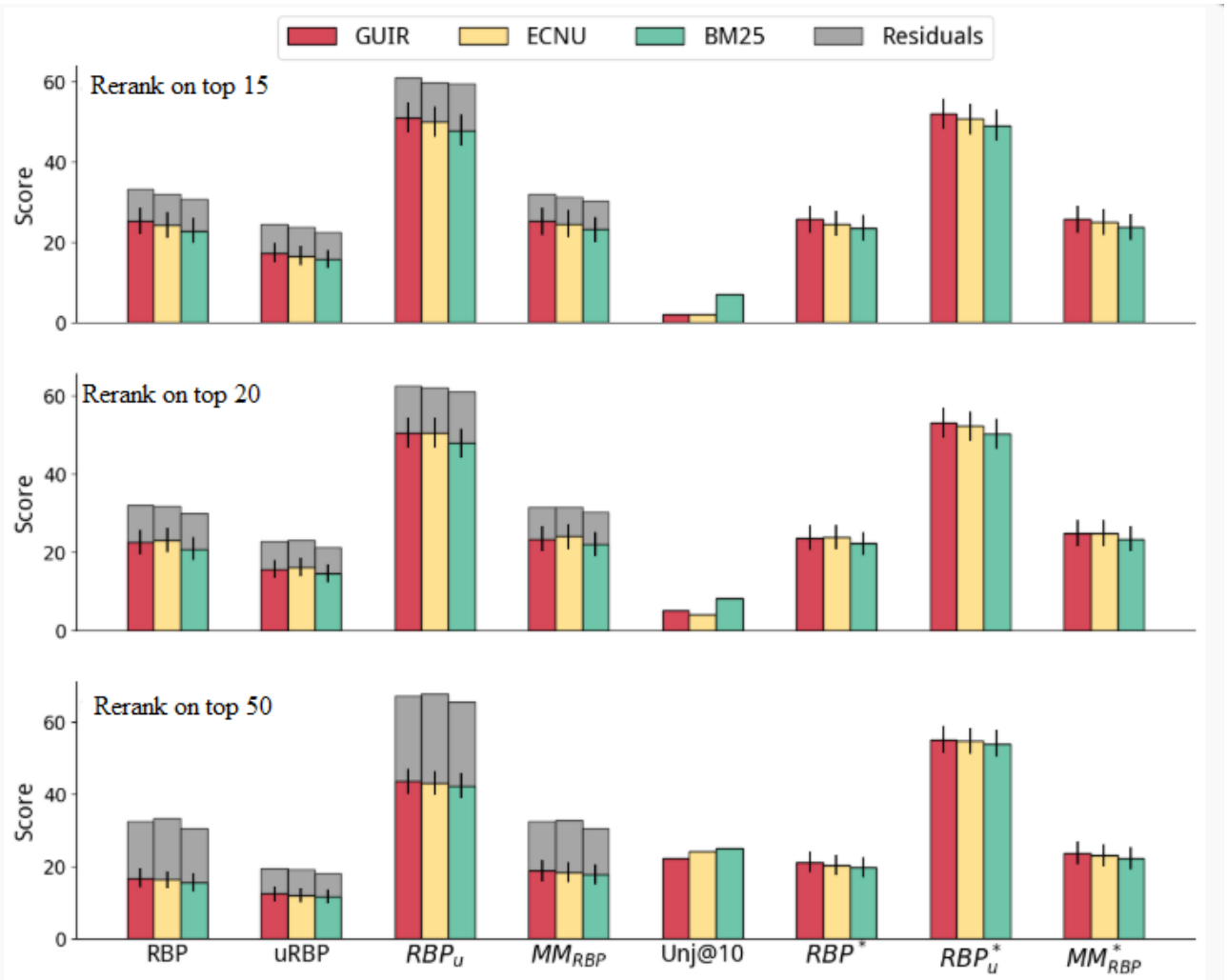


Figure 7. Reranking combining topical relevance (original run) and understandability (eXtreme Gradient Boosting [XGB]) through rank fusion. ECNU: East China Normal University; GUIR: Georgetown University Information Retrieval; MM: multidimensional metric; RBP: rank biased precision.

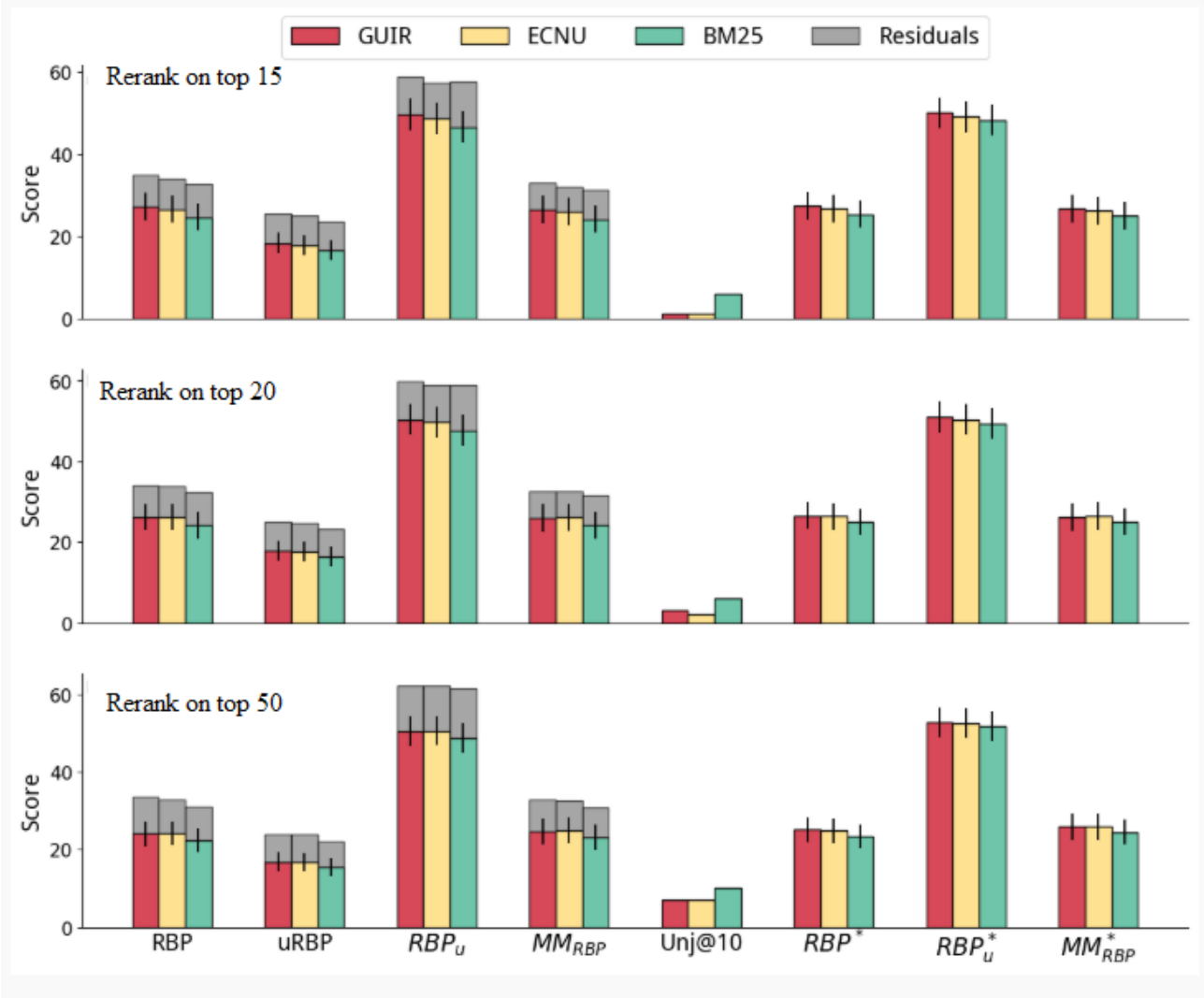
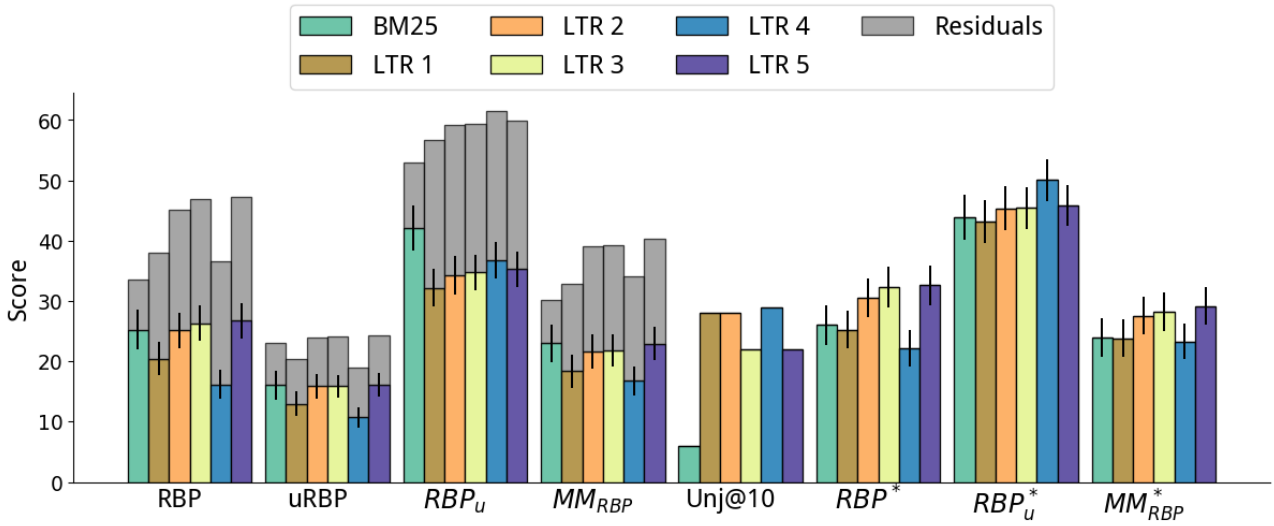


Figure 8. Results of the learning to rank (LTR) method on the Best Match 25 (BM25) baseline. The BM25 baseline (light blue) is shown for direct comparison. MM: multidimensional metric; RBP: rank biased precision.



Simple Reranking

Figure 5 reports the results of reranking methods applied to the runs shown in Figure 4. Reranking was applied based on the DCI score of each document calculated using the preprocessing combination of Boilerpipe and ForcePeriod (best according to Pearson correlation, from Tables 3 and 4). We found that the relevance of the reranked runs (as measured by RBP_r and RBP^*_r) significantly decreased, compared with the original runs, for example, reranking the top 15 search results using DCI made RBP_r decrease from 25.28 to 21.58. However, as expected, these reranked results were significantly more understandable: for the previous example, RBP_u passed from 42.08 to 47.09.

In the experiments, we also studied the influence of the number of documents considered for reranking (cut-off). The top/middle/bottom plots of Figure 5 refer to reranking only the top k=15/20/50 documents from the original runs. The results show that the more documents are considered for reranking, the more degradation in RBP_r effectiveness. Considering understandability only in the evaluation shows mixed results. Similar trends were observed for evaluation measures that consider understandability (RBP and RBP_u), however, with some exceptions. For example, an increase in uRBP was observed when reranking ECNU using the top 50 results.

Note that with the increase of the number of documents considered for reranking, there is an increase in the number of unassessed documents being considered by the evaluation measures. Nevertheless, we note that if unassessed documents are excluded from the evaluation, similar trends are observed, for example, compare findings with those for the condensed measures uRBP*, RBP^*_r, RBP^*_u, and MM^*_{RBP}.

Figure 6 refers to using a machine learning method, XGB regressor (Textbox 9), to estimate understandability. Similarly, when using DCI, as the cut-off increased, for example, from k=15 to k=50, the documents returned were more understandable but less relevant. For the same cut-off value, for example, k=15, the machine learning method used for estimating understandability consistently yielded more understandable results than DCI (higher RBP_u and RBP^*_u).

Overall, statistically significant improvements over the baselines were observed for most configurations and measures.

Rank Fusion

Next, we report the results of automatically combining topical relevance and understandability through rank fusion in Figure 7. We used the XGB method for estimating understandability, as it was the one yielding highest effectiveness for the reranking method. Runs were thus produced by fusing the reranking with XGB and the original run. (Results for DCI are reported in the Multimedia Appendix 5 and confirm the superiority of XGB.)

As for reranking, also for the rank fusion approaches we found that, in general, higher cut-offs were associated to higher effectiveness in terms of understandability measures on one hand, but higher losses in terms of relevance-oriented measures on the other. Overall, results obtained with rank fusion were superior to those obtained with reranking only, although most differences were not statistically significant. Statistically significant improvements over the baselines were instead observed for most configurations and measures.

Learning to Rank

Finally, we analyze the results obtained by the learning to rank methods in Figure 8. Unlike with the previous methods, we did not impose a rank cut-off on learning to rank. Learning to rank was only applied to the BM25 baseline, as we had no access to the IR features for the runs submitted at CLEF (ie, GUIR and ECNU for CLEF 2016). BM25 baseline (Figure 4) is also shown in Figure 8 for an easy and direct comparison.

When considering RBP_r and uRBP, learning to rank exhibited effectiveness that was significantly inferior to that of the GUIR and ECNU baseline runs, although higher than those for the BM25 baseline (for some configurations). The examination of the number of unassessed documents (and the RBP residuals, see Multimedia Appendix 5) revealed that this might have been because measures were affected by the large number of unassessed documents retrieved in the top 10 ranks. For example, the RBP_r residual for learning to rank methods was about double that of the baselines or other approaches (see Multimedia Appendix 5). In fact, among the documents retrieved in the top 10 results by learning to rank, there were 20% (2/10) that were unassessed, compared with an average of 3% (0.3/10) for the other methods (excluding XGB with cutoff 50, which also exhibited high residuals).

We thus should carefully account for unassessed documents through considering the residuals of RBP measures as well as the condensed measures. When this was done, we observed that learning to rank methods overall provided substantial gains over the original runs and other methods (when considering RBP^*_r, RBP^*_u, and MM^*_{RBP}), or large potential gains over these methods (when considering the residuals). Next, we analyzed these results in more detail.

No improvements over the baselines were found for LTR 1, and the high residuals for RBP_r were not matched by other residuals or by considering only assessed documents (see Multimedia Appendix 5). LTR 1 was a simple method that used only IR features and was trained only on topical relevance. Specifically, we devised 24 IR features using the Terrier framework. The score of various retrieval models was extracted from a multifield index composed of title, body, and whole document. Although simple, this is a typical learning to rank setting.

Compared with LTR 1, LTR 2 included the understandability features listed in Textboxes 1-10. This inclusion was as beneficial to the understandability measures as to the relevance measures, with RBP^*_r, RBP^*_u, and MM^*_{RBP} all showing gains over the baselines. LTR 3 obtained similar MM^*_{RBP} values, although with higher effectiveness for relevance measures (RBP^*_r) than for understandability (RBP^*_u).

LTRs 4 and 5 were devised based on a set understandability threshold U=40. Although LTR 4 took into consideration only documents that were easy to read (understandability label≤U),

LTR 5 considered all documents, but boosted the relevance score. LTR 4 reached the highest understandability score for the learning-to-rank approaches (RBP^*_u=50.06), but it failed to retrieve a substantial number of relevant documents (RBP^*_r=2.20). In turn, LTR 5 reached the highest understandability-relevance trade-off (MM^*_{RBP}=29.20). Compared with the BM25 baseline (on which it was based), LTR 5 largely increased both relevance (RBP^*_r from 26.01 to 32.60—a 25% increase, P_{bl}=.003) and understandability (RBP^*_u from 43.89 to 45.87 — a 4% increase, P_{bl}<.001). Note that LTR 5 was also significantly better than the best run submitted to CLEF 2016 for both RBP^*_r (15% increase, P_g=.120) and MM^*_{RBP} (13% increase, P_g=.001).

Discussion

Principal Findings

The empirical experiments suggested the following:

- Machine learning methods based on regression are best suited to estimate the understandability of health Web pages
- Preprocessing does affect effectiveness (both for understandability prediction and document retrieval), although compared with other methods, ML-based methods for understandability estimation are less subjected to variability caused by poor preprocessing
- Learning to rank methods can be specifically trained to promote more understandable search results, whereas still providing an effective trade-off with topical relevance.

Limitations

In this study, we relied on data collected through the CLEF 2015 and CLEF 2016 evaluation efforts to evaluate the effectiveness of methods that estimate the understandability of the Web pages. These assessments were obtained by asking medical experts and practitioners to rate documents; although, they were asked to estimate the understandability of the content as if they were the patients they treat, there might have been noise and imprecisions in the collection mechanism because of the subjectivity of the task. Figure 2 highlights this by showing that the agreement between assessors is relatively low. A better setting might have been to directly recruit health consumers: the task would still have been subjective but would have captured real ratings, rather than inferred or perceived ratings. Despite this, our previous work has shown that no substantial differences were found in the downstream evaluation of retrieval systems, when we acquired understandability assessments from health consumers for a subset of the CLEF 2015 collection [46].

Relevance assessments on the CLEF 2015 and 2016 collections are incomplete [44,14], that is, not all top ranked Web pages retrieved by the investigated methods have an explicit relevance assessment. This is often the case in information retrieval, where the validity of experiments based on incomplete assessments has been thoroughly investigated [88]. Nonetheless, we carefully controlled for the impact that unassessed documents had in our experiments by measuring their number and using measures such as RBP that account for residuals and condensed variants. The residuals analysis has been reported in the appendix.

Conclusions

We have examined approaches to estimate the understandability of health Web pages, including the impact of HTML preprocessing techniques and how to integrate these within retrieval methods to provide more understandable search results for people seeking health information. We found that machine learning methods are better suited than traditionally employed readability measures for assessing the understandability of health-related Web pages and that learning to rank is the most effective strategy to integrate this into retrieval. We also found that HTML and text preprocessing do affect the effectiveness of both understandability estimations and of the retrieval process, although machine learning methods are less sensitive to this issue.

This paper contributes to improving search engines tailored to consumer health search because it thoroughly investigates promises and pitfalls of understandability estimations and their integration into retrieval methods. The paper further highlights which methods and settings should be used to provide better search results to health information seekers. As shown in Figure 1, these methods would clearly improve current health-focused search engines.

The methods investigated here do not provide a fully personalized search, with respect to how much of the health content consumers with different health knowledge might be able to understand. Instead, we focus on making the results understandable by anyone, and thus promote in the search results content that has the highest level of understandability. However, people with a more than average medical knowledge might benefit higher from more specialized content. We leave this personalization aspect, that is, the tailoring of the understandability level of the promoted content with respect to the user's knowledge and abilities, to further work.

Acknowledgments

The authors acknowledge the Technische Universität Wien (TU Wien) University Library for financial support through its Open Access Funding Programme. GZ is the recipient of an Australian Research Council Discovery Early Career Researcher Award (DECRA) Research Fellowship (DE180101579) and a Google Faculty Award. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644753 (KConnect).

Conflicts of Interest

None declared.

Multimedia Appendix 1

The impact of feature sets on the Spearman correlation between the predicted understandability and the ground truth assessed by human assessors in Conference and Labs of the Evaluation Forum (CLEF) eHealth 2015.

[PDF File (Adobe PDF File), 627KB - jmir_v21i1e10986_app1.pdf]

Multimedia Appendix 2

Distribution of Understandability Scores for Conference and Labs of the Evaluation Forum (CLEF) 2016.

[PDF File (Adobe PDF File), 726KB - jmir_v21i1e10986_app2.pdf]

Multimedia Appendix 3

Correlations between understandability estimators and human assessments for Conference and Labs of the Evaluation Forum (CLEF) 2015 and CLEF 2016.

[PDF File (Adobe PDF File), 896KB - jmir_v21i1e10986_app3.pdf]

Multimedia Appendix 4

Correlation results of different readability formulae with human assessments from Conference and Labs of the Evaluation Forum (CLEF) eHealth 2015 and 2016.

[PDF File (Adobe PDF File), 974KB - jmir_v21i1e10986_app4.pdf]

Multimedia Appendix 5

Results obtained by integrating understandability estimations within retrieval methods on Conference and Labs of the Evaluation Forum (CLEF) 2015 and CLEF 2016.

[PDF File (Adobe PDF File), 1MB - jmir_v21i1e10986_app5.pdf]

References

1. Zhang Y, Zhang J, Lease M, Gwizdka J. Multidimensional relevance modeling via psychometrics and crowdsourcing. : ACM; 2014 Presented at: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval; July 06 - 11, 2014; Gold Coast, Queensland, Australia p. 435-444. [doi: 10.1145/2600428.2609577]
2. Fergus TA, Spada MM. Cyberchondria: examining relations with problematic internet use and metacognitive beliefs. Clin Psychol Psychother 2017 Nov;24(6):1322-1330. [doi: 10.1002/cpp.2102] [Medline: 28621035]
3. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
4. White R, Horvitz E. Cyberchondria: Studies of the escalation of medical concerns in Web search. ACM Trans Inf Syst 2009 Nov 01;27(4):1-37. [doi: 10.1145/1629096.1629101]
5. White R. Beliefs and Biases in Web Search. : ACM; 2013 Presented at: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval; July 28 - August 01, 2013; Dublin, Ireland p. 3-12. [doi: 10.1145/2484028.2484053]
6. Graber MA, Roller CM, Kaeble B. Readability levels of patient education material on the World Wide Web. J Fam Pract 1999 Jan;48(1):58-61. [Medline: 9934385]
7. Fitzsimmons PR, Michael B, Hulley J, Scott G. A readability assessment of online Parkinson's disease information. J R Coll Physicians Edinb 2010 Dec;40(4):292-296. [doi: 10.4997/JRCPE.2010.401] [Medline: 21132132]
8. Wiener RC, Wiener-Pla R. Literacy, pregnancy and potential oral health changes: the internet and readability levels. Matern Child Health J 2014 Apr;18(3):657-662. [doi: 10.1007/s10995-013-1290-1] [Medline: 23784613]
9. Patel CR, Cherla DV, Sanghvi S, Baredes S, Eloy JA. Readability assessment of online thyroid surgery patient education materials. Head Neck 2013 Oct;35(10):1421-1425. [doi: 10.1002/hed.23157] [Medline: 22972634]
10. Meillier A, Patel S. Readability of healthcare literature for gastroparesis and evaluation of medical terminology in reading difficulty. Gastroenterology Res 2017 Feb;10(1):1-5 [FREE Full text] [doi: 10.14740/gr746w] [Medline: 28270870]
11. Ellimoottil C, Polcari A, Kadlec A, Gupta G. Readability of websites containing information about prostate cancer treatment options. J Urol 2012 Dec;188(6):2171-2176. [doi: 10.1016/j.juro.2012.07.105] [Medline: 23083852]

12. University of Waterloo. 2016. Cura Te Ipsum: Answering Symptom Queries with Question Intent URL: http://plg2.cs.uwaterloo.ca/~avtyurin/WebQA2016/ [accessed 2018-03-29] [WebCite Cache ID 6yHTeM33k]
13. Boyer C, Dolamic L. Automated detection of HONcode website conformity compared to manual detection: an evaluation. J Med Internet Res 2015 Jun 02;17(6):e135 [FREE Full text] [doi: 10.2196/jmir.3831] [Medline: 26036669]
14. Zuccon G, Palotti J, Goeuriot L, Kelly L, Lupu M, Pecina P. The IR Task at the CLEF eHealth evaluation lab 2016: User-centred Health Information Retrieval. : CEUR-WS; 2016 Sep Presented at: Conference and Labs of the Evaluation Forum; September 5-8, 2016; Évora, Portugal p. 15-27 URL: http://ceur-ws.org/Vol-1609/16090015.pdf
15. Doak C, Doak L, Root J. Teaching Patients with Low Literacy Skills. Philadelphia, Pennsylvania, United States: Lippincott Williams and Wilkins; 1995.
16. Wallace L, Lennon E. American Academy of Family Physicians patient education materials: can patients read them? Fam Med 2004 Sep;36(8):571-574 [FREE Full text] [Medline: 15343418]
17. Davis T, Wolf M. Health literacy: implications for family medicine. Fam Med 2004 Sep;36(8):595-598. [Medline: 15343422]
18. Stossel L, Segar N, Gliatto P, Fallar R, Karani R. Readability of patient education materials available at the point of care. J Gen Intern Med 2012 Sep;27(9):1165-1170 [FREE Full text] [doi: 10.1007/s11606-012-2046-0] [Medline: 22528620]
19. NIH publication. Clear & Simple: Developing Effective Print Materials for Low-literate Readers URL: https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication/ [accessed 2017-09-01] [WebCite Cache ID 6yHTsSTK7]
20. Palotti J, Zuccon G, Hanbury A. The Influence of Pre-processing on the Estimation of Readability of Web Documents. USA: ACM; 2015 Presented at: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management; October 18 - 23, 2015; Melbourne, Australia p. 1763-1766. [doi: 10.1145/2806416.2806613]
21. Palotti J, Goeuriot L, Zuccon G, Hanbury A. Ranking Health Web Pages with Relevance and Understandability. 2016 Presented at: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval; July 17 - 21, 2016; Pisa, Italy p. 965-968. [doi: 10.1145/2911451.2914741]
22. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns 2014 Sep;96(3):395-403. [doi: 10.1016/j.pec.2014.05.027] [Medline: 24973195]
23. Becker SA. A study of web usability for older adults seeking online health resources. ACM Trans Comput-Hum Interact 2004 Dec 01;11(4):387-406. [doi: 10.1145/1035575.1035578]
24. Zheng J, Yu H. Readability formulas and user perceptions of electronic health records difficulty: a corpus study. J Med Internet Res 2017 Mar 02;19(3):e59. [doi: 10.2196/jmir.6962] [Medline: 28254738]
25. Coleman M, Liau TL. A computer readability formula designed for machine scoring. J Appl Psychol 1975;60(2):283-284. [doi: 10.1037/h0076540] [Medline: 28425725]
26. Dale E, Chall J. A formula for predicting readability: instructions. Educational Research Bulletin 1948 Feb 18;27(2):37-54 [FREE Full text]
27. Kincaid J, Fishburne R, Rogers R, Chissom B. National Technical Information Service. Florida, USA: Institute for Simulation and Training, University of Central Florida; 1975 Feb. Derivation of New Readability Formulas for Navy Enlisted Personnel URL: http://www.dtic.mil/dtic/tr/fulltext/u2/a006655.pdf [accessed 2018-12-04] [WebCite Cache ID 74PMHgBnJ]
28. Dubay W. Impact Information. Costa Mesa, CA: Impact Information; 2004 Aug 25. The Principles of Readability URL: http://www.impact-information.com/impactinfo/readability02.pdf [accessed 2018-11-28] [WebCite Cache ID 74GkzsGi0]
29. Liu X, Croft W, Oh P, Hart D. Automatic Recognition of Reading Levels from User Queries. 2004 Presented at: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval; July 25 - 29, 2004; Sheffield, United Kingdom p. 548-549. [doi: 10.1145/1008992.1009114]
30. Collins-Thompson K, Callan J. Predicting reading difficulty with statistical language models. J Am Soc Inf Sci 2005 Nov;56(13):1448-1462. [doi: 10.1002/asi.20243]
31. Heilman M, Collins-Thompson K, Callan J, Eskenazi M. Combining lexical and grammatical features to improve readability measures for first and second language texts. 2007 Presented at: The Conference of the North American Chapter of the Association for Computational Linguistics; April 22-27, 2007; Rochester, New York p. 460-467.
32. Pitler E, Nenkova A. Revisiting Readability: A Unified Framework for Predicting Text Quality. 2008 Presented at: Proceedings of the Conference on Empirical Methods in Natural Language Processing; October 25 - 27, 2008; Honolulu, Hawaii p. 186-195 URL: http://dl.acm.org/citation.cfm?id=1613715.1613742
33. Zeng Q, Kim E, Crowell J, Tse T. A Text Corpora-based Estimation of the Familiarity of Health Terminology. 2005 Presented at: Proceedings of the 6th International conference on Biological and Medical Data Analysis; November 10 - 11, 2005; Aveiro, Portugal p. 184-192 URL: https://dl.acm.org/citation.cfm?id=2146189 [doi: 10.1007/11573067_19]
34. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. J Am Med Inform Assoc 2006;13(1):24-29 [FREE Full text] [doi: 10.1197/jamia.M1761] [Medline: 16221948]
35. McCloskey M, Caramazza A, Basili A. Cognitive mechanisms in number processing and calculation: evidence from dyscalculia. Brain Cogn 1985 Apr;4(2):171-196. [Medline: 2409994]
36. Leroy G, Helmreich S, Cowie J, Miller T, Zheng W. Evaluating online health information: beyond readability formulas. AMIA Annu Symp Proc 2008 Nov 06;2008:394-398 [FREE Full text] [Medline: 18998902]

37. Palotti J, Hanbury A, Muller H. Exploiting Health Related Features to Infer User Expertise in the Medical Domain. 2014 Presented at: Proceedings of WSCD Workshop on Web Search and Data Mining; February 24-28, 2014; New York, USA URL: http://publications.hevs.ch/index.php/publications/show/1632
38. Yan X, Lau RY, Song D, Li X, Ma J. Toward a semantic granularity model for domain-specific information retrieval. ACM Trans Inf Syst 2011 Jul 01;29(3):1-46. [doi: 10.1145/1993036.1993039]
39. Kim H, Goryachev S, Rosemblat G, Browne A, Keselman A, Zeng-Treitler Q. Beyond surface characteristics: a new health text-specific readability measurement. AMIA Annu Symp Proc 2007 Oct 11;2007:418-422 [FREE Full text] [Medline: 18693870]
40. van Doorn J, Odijk D, Roijers D, de Rijke M. Balancing Relevance Criteria through Multi-objective Optimization. 2016 Presented at: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval; July 17 - 21, 2016; Pisa, Italy p. 769-772. [doi: 10.1145/2911451.2914708]
41. Zuccon G, Koopman B. Integrating Understandability in the Evaluation of Consumer Health Search Engines. 2014 Presented at: Medical Information Retrieval (MedIR) Workshop; July 11 2014; Gold Coast, Australia URL: https://eprints.qut.edu.au/72854/
42. Zuccon G. Understandability Biased Evaluation for Information Retrieval. 2016 Mar Presented at: European Conference on Information Retrieval; March 20-23, 2016; Padua, Italy p. 280-292. [doi: 10.1007/978-3-319-30671-1_21]
43. CLEF eHealth. URL: https://sites.google.com/site/clefehealth/ [accessed 2018-04-23] [WebCite Cache ID 6ytaboB6c]
44. Palotti J, Zuccon G, Goeuriot L, Kelly L, Hanbury A, Jones G. ShARe/CLEF eHealth Evaluation Lab 2015, Task 2: User-centred Health Information Retrieval. 2015 Presented at: Conference and Labs of the Evaluation Forum; September 8-11, 2015; Toulouse, France URL: http://ceur-ws.org/Vol-1391/inv-pap9-CR.pdf
45. Lemur. The ClueWeb12 Dataset URL: http://lemurproject.org/clueweb12/ [accessed 2018-04-17] [WebCite Cache ID 6ykSxv4Hp]
46. Palotti J, Zuccon G, Bernhardt J, Hanbury A, Goeuriot L. Assessors agreement: A case study across assessor type, payment levels, query variations and relevance dimensions. : Springer International Publishing; 2016 Presented at: Conference and Labs of the Evaluation Forum; September 5-8, 2016; Évora, Portugal p. 40-53 URL: https://publik.tuwien.ac.at/files/publik_257829.pdf [doi: 10.1007/978-3-319-44564-9_4]
47. Koopman B, Zuccon G. Relevation!: An open source system for information retrieval relevance assessment. 2014 Presented at: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval; July 06 - 11, 2014; Gold Coast, Queensland, Australia p. 1243-1244 URL: https://eprints.qut.edu.au/72102/ [doi: 10.1145/2600428.2611175]
48. Smith EA, Senter RJ. Automated readability index. AMRL TR 1967 May:1-14. [Medline: 5302480]
49. Gunning R. The Technique of Clear Writing. New York City: McGraw Hill Higher Education; 1968.
50. Bjornsson CH. Readability of Newspapers in 11 Languages. Read Res Q 1983;18(4):480-497 [FREE Full text] [doi: 10.2307/747382]
51. Mc Laughlin G. SMOG grading - a new readability formula. Journal of Reading 1969;12(8):639-646 [FREE Full text]
52. Collins-Thompson K. Computational assessment of text readability: a survey of current and future research. ITL 2015 Jan 23;165(2):97-135. [doi: 10.1075/itl.165.2.01col]
53. Pyphen. 2017. URL: http://pyphen.org [accessed 2018-03-29] [WebCite Cache ID 6yHSW2aHz]
54. OpenMedSpel (en-us). 2017. URL: http://extensions.openoffice.org/en/project/openmedspel-en-us [accessed 2018-10-21] [WebCite Cache ID 6yHd3KTZc]
55. Zhou W, Torvik VI, Smalheiser NR. ADAM: another database of abbreviations in MEDLINE. Bioinformatics 2006 Nov 15;22(22):2813-2818. [doi: 10.1093/bioinformatics/btl480] [Medline: 16982707]
56. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
57. Pang C. Minerva Access. 2016. Understanding Exploratory Search in Seeking Health Information URL: https://minerva-access.unimelb.edu.au/handle/11343/115239 [accessed 2018-11-28] [WebCite Cache ID 74GoJu166]
58. Agrafiotes C, Arampatzis A. Augmenting Medical Queries with UMLS Concepts via MetaMap. 2016 Presented at: Proceedings of The Twenty-Fifth Text REtrieval Conference; November 15-18, 2016; Gaithersburg, Maryland, USA URL: https://trec.nist.gov/pubs/trec25/papers/DUTH-CL.pdf
59. Palotti J, Hanbury A, Müller H, Kahn CE. How users search and what they search for in the medical domain. Inf Retrieval J 2015 Oct 24;19(1-2):189-224. [doi: 10.1007/s10791-015-9269-8]
60. Yates A, Goharian N. ADRTrace: Detecting Expected and Unexpected Adverse Drug Reactions from User Reviews on Social Media Sites. 2013 Presented at: European Conference on Information Retrieval; 24-27 March; Moscow, Russia p. 816-819 URL: https://link.springer.com/chapter/10.1007/978-3-642-36973-5_92 [doi: 10.1007/978-3-642-36973-5_92]
61. Pang B, Lee L. Opinion mining and sentiment analysis. Found Trends Inf Ret 2008;2(1–2):1-135. [doi: 10.1561/1500000011]
62. Natural Language Toolkit. 2017. URL: http://www.nltk.org/ [accessed 2017-10-21] [WebCite Cache ID 6yHdLox5S]
63. GNU Aspell. URL: http://aspell.net/ [accessed 2018-03-29] [WebCite Cache ID 6yHdUtryf]

64. Strohman T, Metzler D, Turtle H, Croft W. Indri: A language model-based search engine for complex queries. 2005 Presented at: Proceedings of the International Conference on Intelligent Analysis; May 2 - 3, 2005; Washington, DC p. 2-6 URL: http://ciir.cs.umass.edu/pubfiles/ir-407.pdf
65. Ounis I, Amati G, He B, Macdonald C. Terrier Information Retrieval Platform. 2005 Presented at: European Conference on Information Retrieval; March 21-23, 2005; Santiago de Compostela, Spain p. 517-519. [doi: 10.1007/978-3-540-31865-1_37]
66. Feng L, Jansche M, Huenerfauth M, Elhadad N. A Comparison of Features for Automatic Readability Assessment. 2010 Presented at: Proceedings of the 23rd International Conference on Computational Linguistics; August 23 - 27, 2010; Beijing, China p. 276-284 URL: http://people.dbmi.columbia.edu/noemie/papers/coling10.pdf
67. Barzilay R, Lapata M. Modeling local coherence: an entity-based approach. Comput Linguist 2008 Mar;34(1):1-34. [doi: 10.1162/coli.2008.34.1.1]
68. Beautiful Soup. URL: https://www.crummy.com/software/BeautifulSoup/ [accessed 2018-03-29] [WebCite Cache ID 6yHddZWCi]
69. Elhadad N. Comprehending technical texts: predicting and defining unfamiliar terms. AMIA Annu Symp Proc 2006:239-243 [FREE Full text] [Medline: 17238339]
70. Wu D, Hanauer D, Mei Q, Clark P, An L, Proulx J, et al. Assessing the readability of ClinicalTrials.gov. J Am Med Inform Assoc 2016 Mar;23(2):269-275 [FREE Full text] [doi: 10.1093/jamia/ocv062] [Medline: 26269536]
71. Reddit. URL: https://www.reddit.com [accessed 2018-03-03] [WebCite Cache ID 6yHdMrtgC]
72. Reddit. 2017. Ask a Doctor URL: https://www.reddit.com/r/AskDocs/ [accessed 2018-03-29] [WebCite Cache ID 6yHdhLy3x]
73. PRAW: The Python Reddit API Wrapper. 2017. URL: https://praw.readthedocs.io/en/latest/ [accessed 2018-03-29] [WebCite Cache ID 6yHdm8YI2]
74. Wikimedia. URL: https://dumps.wikimedia.org/enwiki/ [accessed 2018-03-29] [WebCite Cache ID 6yHdZCKxJ]
75. Soldaini L, Cohan A, Yates A, Goharian N, Frieder O. Retrieving Medical Literature for Clinical Decision Support. 2015 Presented at: European Conference on Information Retrieval; 29 March - 2 April, 2015; Vienna, Austria p. 538-549 URL: https://link.springer.com/chapter/10.1007/978-3-319-16354-3_59
76. Roberts K, Simpson M, Demner-Fushman D, Voorhees E, Hersh W. State-of-the-art in biomedical literature retrieval for clinical cases: a survey of the TREC 2014 CDS track. Inf Retrieval J 2015 Jul 18;19(1-2):113-148. [doi: 10.1007/s10791-015-9259-x]
77. Roberts K, Simpson M, Voorhees E, Hersh W. Overview of the TREC 2015 Clinical Decision Support Track. 2015 Presented at: Proceedings of The Twenty-Fourth Text REtrieval Conference; November 17-20, 2015; Gaithersburg, Maryland, USA URL: https://trec.nist.gov/pubs/trec24/papers/Overview-CL.pdf
78. Kohlschütter C, Fankhauser P, Nejdl W. Boilerplate Detection using Shallow Text Features. 2010 Presented at: Proceedings of the third ACM international conference on Web search and data mining; February 04 - 06, 2010; New York, New York, USA p. 441-450. [doi: 10.1145/1718487.1718542]
79. Pomikálek J. Vysokoškolské Kvalifikační Práce. 2011. Removing Boilerplate and Duplicate Content from Web Corpora URL: https://theses.cz/id/nqo9nn/ [accessed 2018-11-29] [WebCite Cache ID 74HmjvbRt]
80. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13 - 17, 2016; San Francisco, California, USA p. 785-794 URL: https://www.kdd.org/kdd2016/papers/files/rfp0697-chenAemb.pdf [doi: 10.1145/2939672.2939785]
81. Cormack G, Clarke C, Buettcher S. Reciprocal rank fusion outperforms condorcet and individual rank learning methods. 2009 Presented at: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval; July 19 - 23, 2009; Boston, MA, USA p. 758-759 URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.2291&rep=rep1&type=pdf [doi: 10.1145/1571941.1572114]
82. Song Y, He Y, Hu Q, He L, Haacke E. ECNU at 2015 eHealth Task 2: User-centred Health Information Retrieval. 2015 Presented at: Conference and Labs of the Evaluation Forum; September 8-11, 2015; Toulouse, France URL: http://ceur-ws.org/Vol-1391/80-CR.pdf
83. Oh H, Jung Y, Kim K. KISTI at CLEF eHealth 2015 Task 2. 2015 Presented at: Conference and Labs of the Evaluation Forum; September 8-11, 2015; Toulouse, France URL: http://ceur-ws.org/Vol-1391/17-CR.pdf
84. Soldaini L, Edman W, Goharian N. Team GU-IRLAB at CLEF eHealth 2016: Task 3. 2016 Presented at: Conference and Labs of the Evaluation Forum; 5-8 September, 2016; Évora, Portugal URL: http://ceur-ws.org/Vol-1609/16090143.pdf
85. Song Y, He Y, Liu H, Wang Y, Hu Q, He L. ECNU at 2016 eHealth Task 3: Patient-centred Information Retrieval. 2016 Presented at: Conference and Labs of the Evaluation Forum; 5-8 September, 2016; Évora, Portugal URL: http://ceur-ws.org/Vol-1609/16090157.pdf
86. Palotti J, Zuccon G, Hanbury A. MM: A new Framework for Multidimensional Evaluation of Search Engines. 2018 Oct Presented at: Proceedings of the 27th ACM International Conference on Information and Knowledge Management; October 22 - 26, 2018; Torino, Italy p. 1699-1702 URL: https://dl.acm.org/citation.cfm?id=3269261 [doi: 10.1145/3269206.3269261]

87. Sakai T. Alternatives to Bpref. 2007 Presented at: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval; July 23 - 27, 2007; Amsterdam, The Netherlands p. 71-78 URL: https://dl.acm.org/citation.cfm?doid=1277741.1277756 [doi: 10.1145/1277741.1277756]
88. Sanderson M. Test collection based evaluation of information retrieval systems. Found Trends Inf Ret 2010;4(4):247-375. [doi: 10.1561/1500000009]

Abbreviations

BM25: Best Match 25
CHV: consumer health vocabulary
CLEF: Conference and Labs of the Evaluation Forum
CLI: Coleman-Liau Index
CMV: consumer medical vocabulary
CRF: components of readability formulae
DCI: Dale-Chall Index
ECNU: East China Normal University
eHealth: electronic health
EMV: expert medical vocabulary
FRE: Flesch Reading Ease
GMV: general medical vocabulary
GUIR: Georgetown University Information Retrieval
HF: HTML features
HON: HealthOnNet
ICD: International Statistical Classification of Diseases and Related Health Problems
KISTI: Korean Institute of Science and Technology Information
LTR: learning to rank
MeSH: Medical Subject Headings
MLC: machine learning classifiers
MLR: machine learning regressors
MM: multidimensional metric
NLF: natural language features
OV: out-of-vocabulary
POS: part of speech
RBP: rank biased precision
RF: readability formulae
SMOG: Simple Measure of Gobbledygook
WFF: word frequency features
XGB: eXtreme Gradient Boosting

Edited by G Eysenbach; submitted 07.05.18; peer-reviewed by C Bond, I Adeleke; comments to author 28.06.18; revised version received 23.08.18; accepted 23.09.18; published 30.01.19.

Please cite as:
Palotti J, Zuccon G, Hanbury A
Consumer Health Search on the Web: Study of Web Page Understandability and Its Integration in Ranking Algorithms
J Med Internet Res 2019;21(1):e10986
URL: http://www.jmir.org/2019/1/e10986/
doi:10.2196/10986
PMID:30698536

Viewpoint

Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line

Anthony R Pisani[1,2], PhD; Nitya Kanuri[3,4], BA (Hons); Bob Filbin[5], MA; Carlos Gallo[6], PhD; Madelyn Gould[7,8], MPH, PhD; Lisa Soleymani Lehmann[9,10], MD, PhD; Robert Levine[11,12], MD; John E Marcotte[13,14], PhD; Brian Pascal[15], JD; David Rousseau[16], MPH; Shairi Turner[5], MD, MPH; Shirley Yen[17], PhD; Megan L Ranney[18], MD, MPH

[1]Department of Psychiatry, University of Rochester Medical Center, University of Rochester, Rochester, NY, United States
[2]Department of Pediatrics, University of Rochester Medical Center, University of Rochester, Rochester, NY, United States
[3]Yale School of Management, Yale University, New Haven, CT, United States
[4]Yale School of Public Health, Yale University, New Haven, CT, United States
[5]Crisis Text Line, New York, NY, United States
[6]Department of Psychiatry and Behavioral Sciences, Northwestern University, Chicago, IL, United States
[7]Department of Psychiatry, Columbia University, New York, NY, United States
[8]Department of Epidemiology, Columbia University, New York, NY, United States
[9]National Center for Ethics in Health Care, Veterans Health Administration, Washington, DC, United States
[10]Harvard Medical School, Harvard University, Boston, MA, United States
[11]Department of Internal Medicine, Yale University, New Haven, CT, United States
[12]Center for Bioethics, Yale University, New Haven, CT, United States
[13]Institute for Social Research, University of Michigan, Ann Arbor, MI, United States
[14]School of Nursing, University of Michigan, Ann Arbor, MI, United States
[15]Center for Internet and Society, Stanford Law School, Palo Alto, CA, United States
[16]Henry J Kaiser Family Foundation, Menlo Park, CA, United States
[17]Department of Psychiatry and Human Behavior, Alpert Medical School, Brown University, Providence, RI, United States
[18]Emergency Digital Health Innovation Program, Department of Emergency Medicine, Alpert Medical School, Brown University, Providence, RI, United States

Corresponding Author:
Anthony R Pisani, PhD
Department of Psychiatry
University of Rochester Medical Center
University of Rochester
300 Crittenden Boulevard Box PSYCH
Rochester, NY, 14642
United States
Phone: 1 585 275 3571
Email: anthony_pisani@urmc.rochester.edu

Abstract

Data sharing between technology companies and academic health researchers has multiple health care, scientific, social, and business benefits. Many companies remain wary about such sharing because of unaddressed concerns about ethics, data security, logistics, and public relations. Without guidance on these issues, few companies are willing to take on the potential work and risks involved in noncommercial data sharing, and the scientific and societal potential of their data goes unrealized. In this paper, we describe the 18-month long pilot of a data-sharing program led by Crisis Text Line (CTL), a not-for-profit technology company that provides a free 24/7 text line for people in crisis. The primary goal of the data-sharing pilot was to design, develop, and implement a rigorous framework of principles and protocols for the safe and ethical sharing of user data. CTL used a stakeholder-based policy process to develop a feasible and ethical data-sharing program. The process comprised forming a data ethics committee; identifying policy challenges and solutions; announcing the program and generating interest; and revising the policy and launching the program. Once the pilot was complete, CTL examined how well the program ran and compared it with other potential program models before putting in place the program that was most suitable for its organizational needs. By drawing

on CTL's experiences, we have created a 3-step set of guidelines for other organizations that wish to develop their own data-sharing program with academic researchers. The guidelines explain how to (1) determine the value and suitability of the data and organization for creating a data-sharing program; (2) decide on an appropriate data sharing and collaboration model; and (3) develop protocols and technical solutions for safe and ethical data sharing and the best organizational structure for implementing the program. An internal evaluation determined that the pilot satisfied CTL's goals of sharing scientific data and protecting client confidentiality. The policy development process also yielded key principles and protocols regarding the ethical challenges involved in data sharing that can be applied by other organizations. Finally, CTL's internal review of the pilot program developed a number of alternative models for sharing data that will suit a range of organizations with different priorities and capabilities. In implementing and studying this pilot program, CTL aimed both to optimize its own future data-sharing programs and to inform similar decisions made by others. Open data programs are both important and feasible to establish. With careful planning and appropriate resources, data sharing between big data companies and academic researchers can advance their shared mission to benefit society and improve lives.

(J Med Internet Res 2019;21(1):e11507) doi:10.2196/11507

KEYWORDS

data sharing; privacy; crisis intervention; text messaging; ethics, business; technology; industry; cooperative behavior; information dissemination

Introduction

Technology companies with large datasets have the potential to fuel discovery in the health sciences and gain valuable insights for their own businesses by sharing their data with academic researchers. Many companies are wary about sharing data with researchers because of concerns about ethics, data security, logistics, and public relations. Previous work published in this journal addresses big data gathered in a health care setting [1,2], but, to date, neither the academic literature nor the technology community has provided guidance for technology companies considering academic data sharing. Without such guidance, few technology companies are willing to take on the potential work and risks involved in noncommercial data sharing, and the scientific and societal potential for their data [3,4] consequently goes unrealized.

In this paper, we describe the 18-month long pilot of a data-sharing program led by Crisis Text Line (CTL), a not-for-profit technology company that provides a free, 24/7 text line for people in crisis in the United States. CTL is the nation's largest provider of crisis interventions via text. From its inception in August 2013 to January 2018, CTL's volunteer crisis counselors have conducted more than 1.7 million conversations with 845,545 unique individuals seeking help for a variety of crises, including suicidal behavior, bullying, self-harm, family conflict, and depression (live data depicting trends across texters can be found at CTL's dedicated tracking website) [5]. The transcripts of these conversations, the metadata they generate (eg, timestamps and area codes), and the postconversation surveys by the crisis counselors (eg, issues encountered and referrals provided) contain rich data about crisis situations. Analyses of these big data have the potential to unearth patterns related both to the needs of youth and adults in crisis and to crisis service delivery, with the attendant possibility of having a wider impact across other fields and organizations that are focused on improving mental health across the world. In addition to sharing insights from its data, CTL has a culture of transparency, continuous learning, and sharing what they have learned in the process of innovation. The information shared in this paper accords with this mission.

CTL wished to find the best ways to share the data they collected with researchers in order to contribute to scientific knowledge, inform mental and public health policy, evaluate and improve the effectiveness of their services, and identify segments of the population that most benefit from CTL services. A pilot data-sharing project was initiated with funding from the Robert Wood Johnson Foundation. The primary goal of the pilot was to design, develop, and implement a rigorous framework of principles that would allow for the safe and ethical sharing of user data, both by CTL and other organizations. CTL's core priority was to develop policies and procedures that would (1) protect the privacy, confidentiality, autonomy, and well-being of its users and (2) protect the reputation, brand, and public trust of the service—issues that are particularly important for CTL, given the profound sensitivity of the crisis situation data. Existing models from other areas of academia provided helpful templates for data-sharing agreements [6] but were insufficient for addressing the full range of needs, risks, and operational challenges of this project. In addition, CTL wanted to ensure that the results of data sharing were actionable and impactful for the larger community.

In this paper, we describe the process of defining core challenges underlying data sharing in technology-academia partnerships; discuss CTL's trial solutions to these challenges; and offer lessons learned that might inform other technology companies' data-sharing partnerships.

The Crisis Text Line Pilot Program

Overview

CTL used a stakeholder-based policy process to develop a feasible and ethical data-sharing program, as described below. The stakeholder process comprised forming a data ethics committee; identifying core challenges to address in establishing a data-sharing program; developing open data-sharing principles and protocols; announcing the program and putting in place the necessary infrastructure; iterative refinement of the protocols and infrastructure; launching the data-sharing program; and evaluation of the results of the pilot. It cost approximately US $900,000 to fund the open data-sharing program pilot. This

funding covered start-up costs; ongoing technical infrastructure costs; 1 full-time open data manager; as well as engineering, data science, and marketing time.

The Data Ethics Committee

CTL convened a panel of academic and technology sector experts to form a data ethics committee. Literature reviews and personal recommendations from the CTL advisory board were used to identify researchers with expertise in data security, research ethics, mobile health intervention, and psychology who would be appropriate committee members. The final data ethics committee had 15 members from 13 institutions and was chaired by CTL's chief data scientist (see Multimedia Appendix 1).

Between January 2016 and April 2017, the data ethics committee met for 4 full group meetings and multiple smaller topic-specific subcommittees via conference call and utilized regular email communications to discuss specific aspects of the data-sharing program. In these meetings, they defined the relevant research ethics concerns, explored options for addressing these concerns, and set criteria for application review. The CTL executive leadership reviewed and approved the policy recommendations put forward by the committee.

Identifying Challenges

The data ethics committee identified 4 core challenges to address in establishing a data-sharing program. These challenges were research ethics, user confidentiality, data security, and threats to the reputation and operations of the service.

Research Ethics

Research on technology-enabled services presents particular opportunities and challenges for protecting the welfare of human subjects [7]; a further set of specific challenges apply in the area of data sharing [8]. For the benefit of organizations that may not be familiar with the principles of research ethics, the 3 core ethical principles highlighted by the data ethics committee were respect for persons (showing regard for individuals' rights to self-determination and privacy), beneficence (doing no harm and maximizing possible benefits), and justice (ensuring equity in access to research and in protection of vulnerable populations) [7,9].

User Confidentiality

The data ethics committee also identified the concept of user data confidentiality (protecting disclosure of identities and information when possible) as critical to a big data-sharing project. Data confidentiality is particularly important for sensitive and potentially stigmatizing data [10], such as that collected by CTL. Large datasets introduce the modern potential challenge of *reidentification* or *deanonymization* of deidentified data provided to third parties: it takes remarkably few pieces of information to identify an individual uniquely, and this task becomes easier the larger and more personalized a dataset is [11]. A further danger is that large datasets may be improperly anonymized [12]. Similarly, blending data from multiple sources also risks deanonymizing the dataset, for the elements in 1 set may fill in the gaps in the other. The CTL data ethics committee was particularly concerned about addressing these potential pitfalls before sharing data.

Data Security

Data security refers to protective measures taken by an organization to prevent unauthorized access to computers, databases, and other confidential information as well as to prevent inadvertent disclosure. The highly sensitive nature of the CTL information as well as the vulnerable nature of the CTL client population mandates very high levels of security precautions when using data. Independent technology companies are generally not subject to the Health Insurance Portability and Accountability Act of 1996 [10] Privacy and Security Rules or the Health Information Technology for Economic and Clinical Health Act of 2009 [13]; nor do they have to develop software and hardware according to the Federal Information Processing Standards 140-2 regulations [14,15]. Nonetheless, maintaining adequate data security and user confidentiality safeguards is critical for client trust and is an ethical imperative. In addition, in applications for institutional review board approval before using external companies' data, academic researchers must demonstrate knowledge of, and adherence to, approved data security best practices for the handling of sensitive and confidential information.

Although all studies should have protocols for avoiding deductive or inadvertent disclosure, highly sensitive data, such as that dealt with by CTL, may require an additional layer of safeguards.

Business Challenges

The 2 primary concerns for a company considering sharing data are reputation and cost. Many technology companies holding data that would be useful to science and society are public-facing companies that rely for business success on their reputation or brand, trust from clients or customers, and the loyalty that follows from it. Negative perceptions can spread quickly through modern media and have the potential to cause significant damage in a relatively short period to a company's image and, by consequence, to its value (as illustrated by recent Facebook scandals).

To mitigate such risks, organizations must calculate and plan for the costs of developing the requisite technical infrastructure and administering a responsible data-sharing program. Technology costs include the development of the data-sharing pipeline and the environment that hosts the data as well as ongoing data hosting fees. Administrative costs comprise principally the costs of staff time and focus. A program manager is required to oversee program marketing, application review, data use agreement execution, and ongoing partner support. Data science and engineering time is required to create custom datasets and offer ongoing technical support.

Although challenges and costs exist, there are many potential benefits that make the investment worthwhile. Collaboration with academic researchers opens up diverse areas of expertise that would be impossible to acquire through internal hiring. In addition, academic-industry partnerships can expand upon the company's original goals to positively impact their population of users. For example, Twitter's data-sharing application programming interface (API) has been used to map restaurant violations [16], HIV infection spread [17], and county-level

heart disease mortality [18]. These uses of the data might not be at the core of the technology company, yet they align with Twitter's original mission to "give everyone the power to create and share ideas and information instantly, without barriers."

Open Data Program Principles and Protocols

The policy development process described above yielded key principles and protocols for addressing the challenges identified by the data ethics committee and CTL leadership. Table 1 provides a summary of these principles and protocols.

Technical and Operational Infrastructure and Program Announcement

After developing these policies and procedures, an open data collaborations (ODC) manager was hired by CTL to develop the requisite technical and operational infrastructure for this study and to ensure adherence to the established policies and procedures. The manager's duties included engaging research teams, coordinating the review of applications, negotiating data use agreements with researchers and their respective universities, developing policies and procedures for sharing custom datasets, and onboarding teams into their data access environments. In addition, the ODC manager, in collaboration with the data ethics committee and the CTL board and staff, developed and iteratively refined the scientific submission and review protocol for applications from academic researchers for data access.

Table 1. Open data program: challenge, principles, and protocols.

Challenges and principles	CTL[a] protocols
Research ethics	
Inform users in an unobtrusive way that anonymized data are shared with select research partners	CTL provides texters with a link to an easy-to-understand Terms of Service[b], including a disclosure of potential future data use, before every crisis conversation
Establish a review process that includes outside academics and ethics experts	An internal CTL team and external ethics committee review applications, with special attention paid to nonmaleficence and justice, texter confidentiality, data security, and social impact
Require human subjects review by academic institution before data sharing	CTL requires each team to procure institutional review board approval
Ensure adequate protection of marginalized groups	CTL reviews research proposals as well as final manuscripts before journal submission for inadvertent stigmatization of marginalized groups (eg, LGBTQ+)
User confidentiality	
Determine which data are released to each team	CTL creates custom datasets for each team, sharing variables on a need-to-know basis for up to 1 year
Protect against release of potentially identifying information	In addition to scrubbing all data for personally identifiable information such as names, addresses, emails, and social media handles, CTL transforms or coarsens any data found to pose a risk to texter confidentiality (eg, university name)
Data security	
Maintain possession of and oversight over data and use	CTL gives each team a virtual machine (VM) hosted on Amazon Web Services and accessed via a virtual private network. All analyses are conducted and stored on the VM with copy/paste and export functionalities disabled
Authorize who can access the data	CTL grants access to university faculty only with demonstration of ethics approval, a signed data use agreement, and a clear data management plan
Require university oversight of, and liability for, researcher behavior when interacting with the data	CTL signs a Data Use Agreement with the lead researcher as well as his or her respective university
Limit the total number of teams to allocate sufficient resources, support, and oversight	CTL limits the number of teams to ≤6 per quarter
Business challenges	
Prioritize research that can benefit users and the service	CTL reviews applications for *value* to texters and crisis community. Projects cannot target for-profit ventures or have plans to monetize research output
Assist with accurate and responsible reporting of results	CTL reviews data output requests and manuscripts before journal submission for accidental breaches of texter confidentiality and accurate contextualization of findings

[a]CTL: Crisis Text Line.

[b]Terms of service: "We have created a formal process for sharing information about conversations with researchers at universities and other institutions. We typically share data with trusted researchers when it will result in insights that create a better experience for our texters. We follow a set of best practices for data sharing based on the University of Michigan's Inter-University Consortium of Social and Political Research, one of the largest open data projects in the U.S., which includes stringent ethical, legal, and security checks. For more details, see our policies for open data collaborations" [19].

CTL staff first reviewed submissions for team competency; proposal feasibility and value; and texter confidentiality, data security, and research ethics. The data ethics committee then reviewed proposals that had been cleared by the CTL staff, specifically looking for red flags related to texter confidentiality, data security, and research ethics, as well as scientific potential. A standardized (iteratively refined) rubric was used for committee scoring of applications.

In accordance with the developed policies, selected research teams received access to a custom dataset hosted on a virtual data enclave on Amazon Web Services (AWS) servers, where all data storage and analyses took place. Finally, to ensure continued compliance with ethical policies, the ODC manager provided ongoing support (eg, onboarding, technical troubleshooting, and data analysis program installation requests) and reviewed data outputs and publications for accidental breaches of texter confidentiality and accurate data contextualization.

Refinement of Protocols and Infrastructure

Between September and December 2016, with ongoing consultation from the ethics committee, CTL piloted and reformulated protocols and created and tested the technical infrastructure, including the virtual data enclave and the ability to create custom datasets. Development and maintenance of this infrastructure were time-intensive and had significant start-up and maintenance costs. The initial development phase took 6 months, with an additional 3 months of user testing and refining of the system to meet researchers' needs. Start-up costs included a call for applications, virtual private network (VPN) hosting, AWS server space, a full-time staff member, and ongoing engineering support. Implementing custom datasets for each research team was also time- and labor-intensive, as it required initial labor to create custom sets and ongoing engineering support for a variety of environments rather than just one.

Pilot Launch

In February 2016, CTL used its website and press releases to issue the first call to researchers for applications to use the data. The press release generated interest from several media outlets, such as FastCompany and BethKanter.org, which ran feature articles. The data ethics committee accepted applications from research teams on a quarterly basis from April 2016 to April 2017. CTL received over 100 applications across 5 quarterly calls for projects. Following both internal and ethics committee reviews, 20 applications were accepted from teams at 18 different universities. Topics included natural language processing (eg, identifying linguistic markers of help seeking among youth who have been abused), correlational studies exploring mental health drivers (eg, correlating weather patterns to service volume and issue prevalence), service use mapping (eg, mapping service use across the state of Montana to visualize unmet needs and inform future resource allocation), and analysis of specific texter populations (eg, use of service by the LGBTQ+ community). Each team received access to a custom dataset housed within a virtual data enclave, as per protocol. The first teams received access to the data in January 2017. The delay between research project review and access to data was approximately 9 months for the first group of approved projects. By January 2017, the timeline for access to data was approximately 3 months. As of the time of this paper's writing, none of the studies have published their analyses, although most are reported to have manuscripts in development.

Refining the Program

As the pilot program was nearing completion, CTL conducted an internal evaluation of the program. CTL used the following criteria to evaluate the success of the program: value (to science and the organization); ethical principles and policies; the ability to share data while maintaining user confidentiality; the ability to provide secure access while maintaining control of data; and the ability to support the program with adequate financial, human, and infrastructure resources. CTL was satisfied that the data-sharing program had value, both to science and to the company, as it prompted greater understanding of data and provided opportunities to share important insights with a broader community. CTL and members of the data ethics committee were also satisfied that the program met its ethical ideals in both principle and practice, that the confidentiality of texters was adequately protected, and that it was feasible to share custom datasets with each research team.

The application process and the data-sharing processes gave CTL confidence that teams were competent, collaborative, and likely to bring value to CTL users and the community at large. These processes provided means to mitigate the user confidentiality and data security risks described above by combining traditional data-sharing processes (eg, use of a VPN and sharing of minimum necessary data for a project) with innovative, technology-specific solutions (eg, a custom-built virtual desktop with safeguards to prevent the copying of data to local machines). However, the financial, infrastructure, and human capital requirements for maintaining a safe, stand-alone open data program were identified as challenging, and the cost and effort of supporting the pilot program were higher than anticipated.

Although start-up funding covered the development of a custom system, maintenance proved to be just as expensive as, if not more than, initial development, as is the case with most rapidly evolving technologies. A related issue is that data hosting is expensive, particularly if each team requires a custom server with custom data. Hosting data for 1 team on AWS servers, for example, costs approximately US $500 a month. Given that the pace of technological change will only increase, CTL identified the difficulties and costs involved in the necessary ongoing auditing and iteration of technical infrastructure as an additional future financial burden. Ultimately, the resources needed to develop and effectively run the data center and related technology significantly outstripped expectations.

In addition to cost, the resources required to oversee the program and provide effective support for research teams were underestimated. In addition to setting teams up with custom data, CTL had to provide ongoing support to add project-specific data analysis software to the virtual server, to clarify aberrations in the data, to troubleshoot bugs, and to review and approve research outputs.

Ethical research collaborations also require communication at every step (data access, custom requests, and output review), and the frequency and difficulty of such communications are amplified when working with researchers from outside the organization. CTL found that the use of human resources and the diversion of focus were higher than originally anticipated. Finally, although free data are desirable to researchers, and therefore drove a high number of applications in this study, researchers found it challenging to focus their time on an unfunded study among their other academic priorities and funded studies. As a result, the pace of research was often slow, with many researchers conducting analyses during weekends and personal time.

On the basis of their internal review, CTL determined that the data-sharing pilot showed promise, but discussion arose about alternatives that might better suit their specific organizational framework and available resources. Resource allocation, both in terms of funding and personnel, is challenging if data sharing is neither part of an organization's core competency nor a core business objective.

Alternative Research Models

In response to the perceived strengths and weaknesses of the initial data-sharing pilot, CTL considered different models for future collaboration with researchers, including alternatives for structuring access to data and for organizing the management of data-sharing projects. The approach pursued in the pilot was designed to maximize the openness of the program by accepting applications from all interested researchers and by approving a large number of projects. This had the advantage of bringing the widest possible range of scholarly perspectives to bear on the data and matched well with CTL's history of working with crowdsourced capabilities and their commitment to open access. However, it also proved to be a costly model, requiring significant start-up funds and the creation of the necessary technical infrastructure as well as the ongoing provision of tools to allow secure remote access to the data by researchers working across a large number of locations. The required degree of institutional focus was also high, with the program needing a considerable commitment of manpower and attention that, at times, strained CTL's ability to support its primary mission.

Two alternative, less resource-intensive models were considered. The first involved the funding of a smaller number of resident researchers, who would apply for 3 to 6 months residential fellowships, during which time they would have on-site access to data. This approach would eliminate the cost of developing and maintaining a data center, reduce expenditure of manpower and focus through easing communication and collaboration, and minimize data security concerns. However, this approach would also reduce the number of teams able to work with the data and, thus, the quantity of outputs from the program. In addition, it would restrict access on a geographical basis, limiting participation to those researchers who are physically close enough to take part in the program.

The second alternative involved collaborating closely on an ongoing basis with a select group of trusted research partners. This approach has the advantage of heavily streamlining communications and minimizing distractions from other tasks, while also increasing the organization's voice both in guiding research questions and in the dissemination of research findings. It also provides financial savings in the maintenance of data-sharing facilities. However, working with a group of trusted partners minimizes the open access nature of the research and, thus, reduces the scope and quantity of research projects.

Ultimately, CTL decided to pursue both alternative strategies concurrently, with each strategy offering a different way of satisfying the underlying principles developed in the first phase of the pilot, often through the implementation of the existing protocols. First, CTL decided to work with very close, highly regarded research partners who would effectively become an extension of the CTL team in the long term. Second, they created an in-house fellowship to allow academic researchers to conduct research via short-term on-site residencies at CTL's primary location with extremely close oversight. These 2 approaches significantly reduced costs, afforded greater data security, and hold out the promise of increasing opportunities to connect research to outcomes of direct value to the organization.

Supplementary Management Tools

Finally, CTL considered 2 different ways of reducing costs and resource usage for data sharing, through the adoption of alternative management models. The first involved using a third-party vendor to manage data warehousing when working with external research partners. This approach reduces the financial and personnel costs of creating and maintaining a data center, increases data security by exploiting the core competencies of the third party, and removes a major attention overhead, thus allowing for an increased focus on leveraging and communicating research outputs. However, because of the necessity of handing over control of highly sensitive data, the organization needs to work with a closely vetted partner. It may also be the case that although small- and medium-sized organizations will reduce their costs by outsourcing the data management element of the program, in very large organizations, it may be possible to scale internal data management solutions to a level that makes the internal option more cost-effective. The revised program adopted by CTL ultimately opted for on-site data access for both in-house research fellows and trusted partners, as the trusted partners were located within easy traveling distance of CTL's primary site. However, third-party management of remote data access would be compatible with either the original program or a trusted partners program working with partner institutions that are geographically distant.

A second possible option for more efficient management of data-sharing projects is the creation of a separate company dedicated to sharing data. This approach reduces the cost and level of distraction for the parent organization and opens up the possibility of financing data-sharing programs through a different funding model, such as charging an administrative fee for data access, which researchers could build into grant proposals. However, the parent company will still need to provide oversight and consultation for the spin-off company because staff in the parent company will have a unique view into the data and their meaning. CTL ultimately decided that their needs could be met through the other revisions made to

their program and, thus, opted not to pursue the spin-off company structure.

Guidelines for Organizations Considering Data Sharing

CTL created a trial data-sharing program both to explore how their collection of data could generate new benefits for their users and to assist other helping organizations through providing access to CTL's data and feedback as well as to the lessons learned in the innovation process. Although it proved feasible and ethical to create the pilot program, the cost of sustaining this particular model was ultimately too high in the light of other organizational priorities. However, this set of principles provided a secure foundation on which CTL was able to iterate further in the development of new data-sharing programs. The principles also offer a framework that other organizations can now adapt to create safe and ethical programs of their own, allowing the broader sharing of data while protecting the privacy of users.

On the basis of our experiences, we suggest that organizations considering data sharing with academic researchers pursue the approach outlined, which has been reverse engineered from the questions studied by CTL during their evaluation phase.

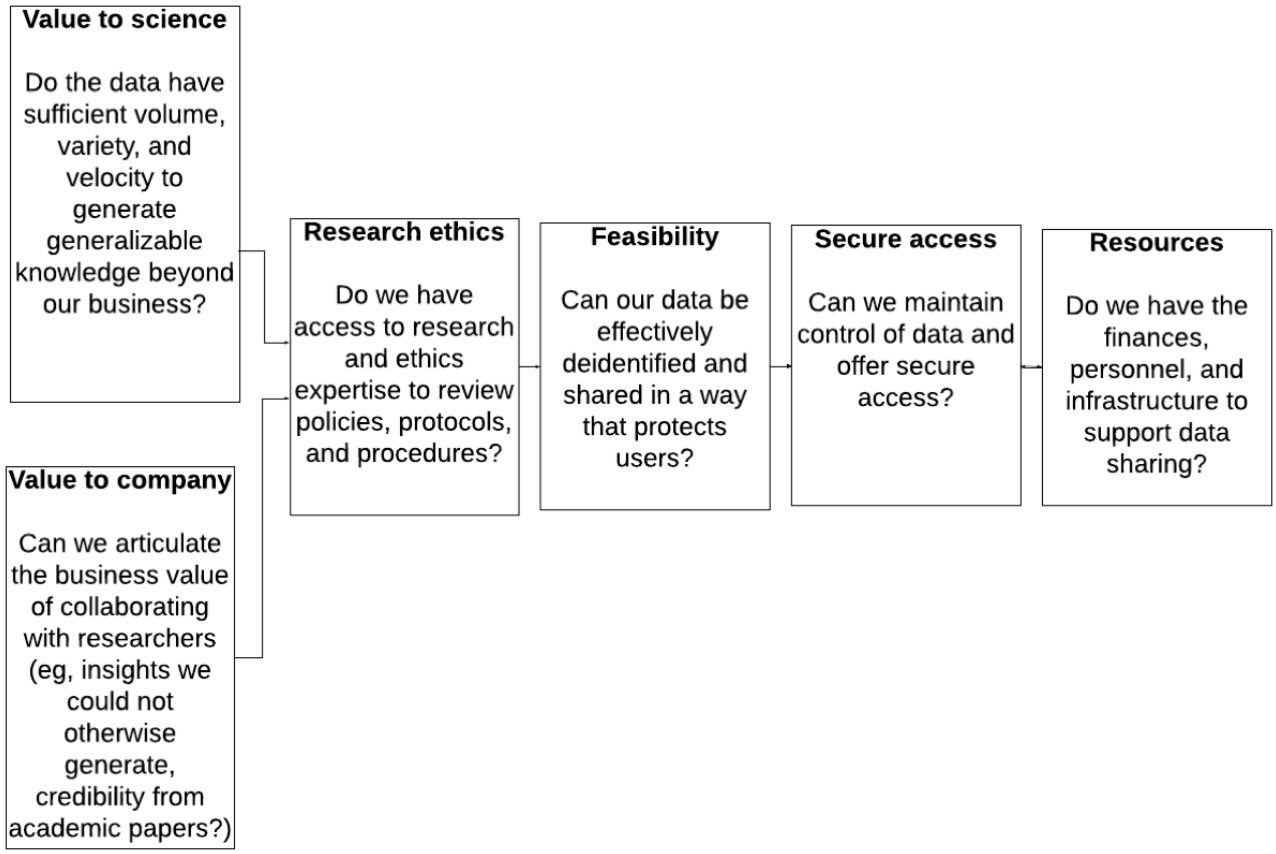
Determine the Value and Suitability of the Data and Organization

The first step for organizations considering academic data sharing is to establish (1) whether the data they have collected are, in principle, valuable to science and (2) whether academic study of the data will further the goals of the organization and/or, in the case of profit-oriented companies, whether study of the data will add value to the business. Once it is determined that suitable data are available for sharing, the following 4 questions need to be answered to determine whether an ethically rigorous and practically feasible program can be established:

- Do we have access to research and ethics expertise to review policies, protocols, and proposals?
- Are our data of such a type that they can be deidentified effectively and shared with researchers in a manner that protects user confidentiality?
- Can we offer secure portals for accessing data?
- Do we have the financial resources, the human capital, and the physical and digital infrastructure to support ethical sharing of data without undermining other organizational priorities?

If all of these questions yield affirmative answers, the organization can then move on with confidence to developing the policies that will guide their particular data-sharing program (see Figure 1 for a summary).

Figure 1. Key questions for organizations considering data sharing.



Decide on an Appropriate Data Sharing and Collaboration Model

In developing its data-sharing program, CTL engaged in a rigorous process of review that examined the pros and cons of the initial pilot program and then studied 2 further potential program structures. The 3 different models considered in this study, along with their advantages and disadvantages, are provided in Table 2. Organizations considering data sharing can draw on CTL's experiences to help identify the model that is most appropriate for them.

Table 2. Research partnership models.

	Open data-sharing program (pilot)	Resident researchers	Selective research partnership
Summary	Open application process for multiple teams to access data and conduct diverse studies at a distance	Researchers apply for 3 to 6 months on-site residency with access to data via computers maintained by the organization	Collaborate closely with a select few trusted research partners on a long-term ongoing basis
Pros	Maximizes variety and quantity of research projects	Eliminates cost of developing data center; eases communication and collaboration; and reduces data security concerns	Increases the organization's voice in guiding research questions and operating principles and increases control over dissemination of research findings
Cons	Most costly option, requiring both start-up and maintenance funding and personnel	Geographic limitation to research collaborators and fewer teams at once	More limited scope of research

Develop Protocols and Technical Solutions for Safe and Ethical Data Sharing and the Best Organizational Structure for Implementing the Program

Once an appropriate data-sharing model has been selected, it will be necessary to draw on the available research and ethics expertise to develop policies and protocols that match the principles of the organization, and then to review research proposals. The pilot conducted by CTL has provided a transferable set of principles (see Table 1, column 1) and a prototype for protocols (see Table 1, column 2). Each organization will have to determine how to apply these principles in their own particular niche (see Table 3).

Table 3. Data management models.

	Internal data management	Third-party data management
Summary	The organization manages data warehousing and access solutions	A third-party vendor manages the data warehousing for an external partner
Pros	Provides maximum control over data security and increases responsiveness to needs of the organization and researchers	Reduces technical costs of starting and maintaining a data center; increases data security, given the third party's core competencies; and enables focus on leveraging and communicating research outputs
Cons	Significant expenses involved in starting and maintaining a data center and draws focus away from organization's core competencies	Organization loses some control over data, therefore must work with a vetted partner

It is not a small thing to institute a program of this sort. Organizations that decide to create a data-sharing program will need to decide how much energy and resources they can reasonably devote to it and assess how the implementation of the project will affect their overall mission. If creating the data solution, or even the overall oversight structure, will become too overwhelming in terms of resources or focus, organizations might consider outsourcing data management or even spinning off a separate organization (for-profit or nonprofit) to reduce these burdens.

Conclusions

The implementation and evaluation of the CTL open data collaboration pilot provides a planning model for *big data* technology companies interested in collaborating with academics. This extends previous work on human subject concerns [7] and on big data in health care [1,2] by focusing specifically on the ethical and practical considerations for data sharing and collaboration in prevention research. We identified key principles in the areas of research ethics, user confidentiality, data security, and business challenges and then developed innovative protocols to ensure that these principles governed the manner in which data were shared. The lessons learned in this process, and in the evaluation of the pilot model, have been distilled into a set of guidelines that can be used by other organizations considering their own academic data-sharing programs.

Big data are generated all the time from a myriad of sources, each with its own potential value to science and its own ethical challenges. We believe that the principles and protocols offered here can be applied and adapted to a broad variety of contexts. However, some limitations should be noted. First, the data collected by CTL are, for the most part, of such a type that users are aware of what they are sharing. In other contexts, such as data collection from passive sources, eg, Global Positioning System sensors, researchers must pay additional attention to the right to privacy of users as well as to obtaining informed consent. Second, although the volume of data generated by CTL is extremely high, some technology companies have datasets that are even larger and more complex. Organizations wishing

to share these larger datasets will face additional challenges. Potential research teams will need to be vetted to ensure that they have the technical and data science capabilities to make the best possible use of these datasets. In parallel, the organization will also need to ensure that data are presented in a format that will be as easy as possible for researchers to handle.

On the basis of both the successes and challenges of the CTL data-sharing pilot, we urge organizations that are considering sharing data with academic researchers to evaluate their preparedness carefully and determine which models are most suitable for their specific institutional needs. In addition to the type of internally managed open data-sharing program trialed by CTL, technology companies may also wish to consider other research partnership models, such as those summarized in Table 2, and other approaches to data management, such as those in Table 3. Key variables to consider include the company's objectives for data sharing, the degree of control they wish to maintain, and their available financial and staffing resources.

Open data programs are important and feasible to establish. Companies embarking on such projects should be aware of the significant commitments and responsibilities involved in the sharing of data. With careful planning and appropriate resources, data sharing between big data companies and academic researchers can advance their shared mission to benefit society and improve lives.

Acknowledgments

The authors would like to dedicate this paper to the memory of Dr Shalon M Irving. The authors gratefully acknowledge support from the Robert Wood Johnson Foundation. The views expressed do not necessarily reflect the views of the Foundation. The authors would also like to acknowledge the support of the American Foundation for Suicide Prevention grant *Youth Who Text a Crisis Line: Understanding Needs and Help-seeking,* SRG-0-110-15, and the National Institute of Mental Health grants K23MH101449 and K23MH098566. Additional support was provided by grant DA027828 from the National Institute on Drug Abuse, awarded to C Hendricks Brown. The authors thank Paul Scade for assistance with manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Crisis Text Line data ethics committee members.

[PDF File (Adobe PDF File), 25KB - jmir_v21i1e11507_app1.pdf]

References

1. Badawi O, Brennan T, Celi LA, Feng M, Ghassemi M, Ippolito A, MIT Critical Data Conference 2014 Organizing Committee. Making big data useful for health care: a summary of the inaugural MIT critical data conference. JMIR Med Inform 2014;2(2):e22. [doi: 10.2196/medinform.3447] [Medline: 25600172]
2. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and opportunities of big data in health care: a systematic review. JMIR Med Inform 2016 Nov 21;4(4):e38 [FREE Full text] [doi: 10.2196/medinform.5359] [Medline: 27872036]
3. Smith M, Patil DJ, Munoz C. Obama White House Archives. 2016. Big Risks, Big Opportunities: the Intersection of Big Data and Civil Rights 2016 URL: https://obamawhitehouse.archives.gov/blog/2016/05/04/big-risks-big-opportunities-intersection-big-data-and-civil-rights [accessed 2018-07-06] [WebCite Cache ID 70hn3nBao]
4. Jagodnik KM, Koplev S, Jenkins SL, Ohno-Machado L, Paten B, Schurer SC, et al. Developing a framework for digital objects in the Big Data to Knowledge (BD2K) commons: Report from the Commons Framework Pilots workshop. J Biomed Inform 2017 Dec;71:49-57 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.006] [Medline: 28501646]
5. Crisis Trends. URL: https://crisistrends.org/ [accessed 2018-12-03] [WebCite Cache ID 74OBEY64b]
6. Kanous A, Brock E. Deep Blue Library, University of Michigan. 2015. Model Data Sharing Agreement URL: https://deepblue.lib.umich.edu/handle/2027.42/136146 [accessed 2018-12-02] [WebCite Cache ID 74MS0KAf6]
7. Pisani AR, Wyman PA, Mohr DC, Perrino T, Gallo C, Villamar J, et al. Human subjects protection and technology in prevention science: selected opportunities and challenges. Prev Sci 2016 Aug;17(6):765-778 [FREE Full text] [doi: 10.1007/s11121-016-0664-1] [Medline: 27220838]
8. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. PLoS One 2011;6(6):e21101 [FREE Full text] [doi: 10.1371/journal.pone.0021101] [Medline: 21738610]
9. US Department of Health, Education, and Welfare. 1979. The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research URL: https://videocast.nih.gov/pdf/ohrp_appendix_belmont_report_vol_2.pdf [accessed 2018-12-02] [WebCite Cache ID 74MSiu8dg]
10. United States Government Publishing Office. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) URL: https://www.gpo.gov/fdsys/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf [accessed 2018-07-06] [WebCite Cache ID 70hnBsZpa]

11. Golle P. Revisiting the Uniqueness of Simple Demographics in the US Population. 2006 Presented at: 5th ACM workshop on Privacy in electronic society; November 03, 2006; Alexandria, Virginia, USA URL: https://www.privacylives.com/wp-content/uploads/2010/01/golle-reidentification-deanonymization-2006.pdf
12. Narayanan A, Shmatikov V. Robust De-anonymization of Large Sparse Datasets. : IEEE; 2008 Presented at: 2008 IEEE Symposium on Security and Privacy (sp 2008); May 18-22, 2008; Oakland, CA, USA URL: https://ieeexplore.ieee.org/document/4531148
13. Blumenthal D. Implementation of the federal health information technology initiative. N Engl J Med 2011 Dec 22;365(25):2426-2431 [FREE Full text] [doi: 10.1056/NEJMsr1112158] [Medline: 22187990]
14. National Institute of Standards and Technology. 2014. Framework for Improving Critical Infrastructure Cybersecurity URL: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf [accessed 2018-12-02] [WebCite Cache ID 74MTmxmld]
15. EDUCAUSE. 2015. Security Professionals Conference: Security Professionals Go “Back to the Future...” URL: https://events.educause.edu/security-professionals-conference/2015 [accessed 2018-12-02] [WebCite Cache ID 74MU4SdMh]
16. Harris JK, Mansour R, Choucair B, Olson J, Nissen C, Bhatt J, Centers for Disease Control and Prevention. Health department use of social media to identify foodborne illness - Chicago, Illinois, 2013-2014. MMWR Morb Mortal Wkly Rep 2014 Aug 15;63(32):681-685 [FREE Full text] [Medline: 25121710]
17. Young SD, Rivers C, Lewis B. Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. Prev Med 2014 Jun;63:112-115 [FREE Full text] [doi: 10.1016/j.ypmed.2014.01.024] [Medline: 24513169]
18. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]
19. Crisis Text Line. Terms of Service & Privacy Policy URL: https://web.archive.org/web/20170612220236/https://www.crisistextline.org/privacy [accessed 2018-12-03] [WebCite Cache ID 74OBQUcqB]

Abbreviations

API: application programming interface
AWS: Amazon Web Services
CTL: Crisis Text Line
ODC: open data collaborations
VPN: virtual private network

Edited by G Eysenbach; submitted 13.07.18; peer-reviewed by Z Predmore, B Brumen; comments to author 05.08.18; revised version received 31.10.18; accepted 13.11.18; published 17.01.19.

Please cite as:
Pisani AR, Kanuri N, Filbin B, Gallo C, Gould M, Lehmann LS, Levine R, Marcotte JE, Pascal B, Rousseau D, Turner S, Yen S, Ranney ML
Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line
J Med Internet Res 2019;21(1):e11507
URL: https://www.jmir.org/2019/1/e11507/
doi:10.2196/11507
PMID:30664452

Original Paper

Leading by Example: Web-Based Sexual Health Influencers Among Men Who Have Sex With Men Have Higher HIV and Syphilis Testing Rates in China

Dan Wu[1,2,3,4], MBBS, PhD; Weiming Tang[1,2,3], PhD; Haidong Lu[1,5], MSc; Tiange P Zhang[1,6], MBBS; Bolin Cao[7], PhD; Jason J Ong[4,8], MBBS, PhD; Amy Lee[1,2], BS; Chuncheng Liu[9], BS; Wenting Huang[1,2], MSc; Rong Fu[2]; Katherine Li[10], MD; Stephen W Pan[11], PhD; Ye Zhang[1,2,12], MPH; Hongyun Fu[13], PhD; Chongyi Wei[14], PhD; Joseph D Tucker[1,2,4], MD, PhD

[1]University North Carolina at Chapel Hill, Project-China, Guangzhou, China
[2]Social Entrepreneurship to Spur Health Global, Guangzhou, China
[3]Dermatology Hospital of Southern Medical University, Guangzhou, China
[4]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom
[5]Department of Epidemiology, University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, NC, United States
[6]Stritch School of Medicine, Loyola University Chicago, Maywood, IL, United States
[7]School of Media and Communication, Shenzhen University, Shenzhen, China
[8]Central Clinical School, Monash University, Monash, Australia
[9]Department of Sociology, University of California, San Diego, CA, United States
[10]Weill Cornell Medical College, New York, NY, United States
[11]Department of Health and Environmental Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, China
[12]Kirby Institute, University of New South Wales, Sydney, Australia
[13]Eastern Virginia Medical School, Norfolk, VA, United States
[14]Department of Social and Behavior Health Sciences, Rutgers School of Public Health, Piscataway, NJ, United States

Corresponding Author:
Dan Wu, MBBS, PhD
University North Carolina at Chapel Hill, Project-China
No 2 Lujing Road
Yuexiu District
Guangzhou, 510091
China
Phone: 86 16602065374
Email: wudan@seshglobal.org

Abstract

Background: The spread of healthy behaviors through social networks may be accelerated by influential individuals. Previous studies have used lay health influencers to prevent sexually transmitted infections (STIs) among internet-using men who have sex with men (MSM). However, there is a lack of understanding of the characteristics of this key subset of MSM.

Objective: This study aimed to examine sociodemographic characteristics, HIV and syphilis testing, and sexual behaviors of Web-based MSM sexual health influencers (SHIs) in China, defined as individuals with relatively stronger influence on spreading HIV and STI information online.

Methods: A Web-based survey of MSM was conducted in August 2017 as a final follow-up of a randomized controlled trial promoting HIV testing in 8 Chinese cities. Men were recruited through a gay social networking mobile phone app and were included if they were born biologically male, aged 16 years and above, ever had sex with another man, and HIV negative or with unknown HIV status. Information regarding sociodemographic characteristics, sexual behaviors, and HIV and syphilis testing was obtained. We assessed men's Web-based sexual health influence using a standardized 6-item opinion leadership scale focused on HIV and STI information. Influencers were defined as those whose mean score ranked within the top 13% (a higher score means greater influence). We used multivariable linear and logistic regression models to measure Web-based sexual health

influence's association with HIV and syphilis testing, controlling for intervention trial effects, age, education, income, and marital status.

Results: Overall, 1031 men completed the survey. Most men were younger than 30 years (819/1031, 79.43%) and had at least college education (667/1031, 64.69%). Influencers were more likely to get tested for HIV (73/132, 55.3% vs 337/899, 37.5%; *P*<.001) and syphilis (35/132, 26.5% vs 137/899, 15.2%; *P*=.001) in the last 3 months compared with noninfluencers. There were no significant differences in condomless sex with male partners (26/132, 19.7% vs 203/899, 22.6%; *P*=.46), mean number of male sex partners (1.32 vs 1.11; *P*=.16) in the last 3 months, and mainly meeting male sex partners online in the last 12 months (97/132, 73.5% vs 669/899, 74.4%; *P*=.82) between influencers and noninfluencers. Regression analyses showed that influencers had higher odds of HIV testing (adjusted odds ratio, AOR 2.16, 95% CI 1.48-3.17) and syphilis testing (AOR 1.99, 95% CI 1.28-3.10) in the last 3 months.

Conclusions: We identified Web-based SHIs who might be more likely to help promote healthy HIV and syphilis testing behaviors through MSM populations. Leveraging existing influencers may help improve HIV and syphilis testing among their networks.

(J Med Internet Res 2019;21(1):e10171) doi:10.2196/10171

KEYWORDS

health promotion; peer influence; internet; social networks; social media; HIV; syphilis; men who have sex with men; China

Introduction

Background

Men who have sex with men (MSM) continue to be disproportionately affected by HIV globally [1]. In Asia, MSM account for 18% of all newly identified HIV diagnosis, and the rising prevalence has been documented in China [2-6]. The national HIV prevalence among MSM in China in 2015 was 8.0%, 200 times higher than 0.04% in China's general population [7]. The low HIV testing rate continues to contribute to the spread of HIV among MSM. Although China has significantly expanded its efforts to control HIV transmission [8], systematic reviews suggest that the HIV testing rate among the Chinese MSM remains low [9-11]. Barriers to HIV testing include the lack of MSM community engagement, hesitancy to access facility-based services, and low level of trust toward facility-based services [12].

Influential individuals facilitate the spread of certain behaviors within a population, as demonstrated in both diffusion research [13,14] and social network research [15-17]. The social diffusion theory suggests that behavior change in a population can be initiated and diffused to others if a behavior is visibly endorsed by enough natural and influential opinion leaders within the population [14]. These influential individuals informed the concept of *popular opinion leaders (POLs)* and the development of POL-based HIV interventions that successfully identified and trained popular individuals to spread HIV prevention messages to peers [18,19]. Social networks can amplify the spread of behavior through interpersonal ties and have been used by public health interventions to promote a range of health behaviors [20-22]. The targeting of influential individuals also makes social network interventions more effective and efficient, possibly because of optimal properties associated with an influential individual's network structure [17]. This body of evidence indicates that behavioral change in social networks may be accelerated through influential individuals.

In the internet era, MSM increasingly turn to Web-based networks and social media to look for sexual health information [23-25]. Popular MSM who frequently share sexual health information with their friends or social media followers may influence the health behaviors of people among their networks. However, understanding of such influential individuals in Web-based MSM networks remains incomplete. First, although social media is more available to MSM as a way to disseminate health knowledge [26], its use among Web-based influential individuals has not been fully explored [19]. Second, unlike POL interventions where the behaviors spread by influential individuals are purposefully established by POL-targeted training, Web-based networks also facilitate the spread of behaviors that are naturally endorsed by influential individuals themselves [15]. However, the sexual risk behaviors, HIV and syphilis testing behaviors of Web-based influential individuals have not been thoroughly described, creating uncertainty to the degree of these individuals' positive influence.

Knowledge of both the use of social media and the set of health behaviors endorsed by Web-based influential individuals can inform the development of MSM-led, network-based interventions. The high rates of internet and mobile phone usage in China particularly provide a strong foundation for network-based interventions on social media [27-29].

Objectives

In this study, we aimed to examine *Web-based sexual health influencers (SHIs)*, defined as individuals with strong influence on spreading HIV and sexually transmitted infection (STI) information online, among Chinese MSM to describe their social-behavioral characteristics, including sexual behaviors, HIV and syphilis testing, social media engagement, community engagement, and HIV-relevant psychological profiles such as anticipated HIV stigma and HIV testing self-efficacy.

Methods

Study Population

An online survey of 1031 MSM was conducted in August 2017 as a final follow-up of a step-wedged randomized controlled trial to improve HIV testing rates in 8 Chinese cities

(Guangzhou, Jiangmen, Zhuhai, and Shenzhen in Guangdong province and Yantai, Jinan, Qingdao, and Jining in Shandong province). The cities of each province were randomly assigned the order of intervention and paired into 4 groups accordingly. The methods have been described in detail elsewhere [30]. Men were recruited through a gay social networking mobile phone app, Blued, by sending a survey invitation to registered users in the 8 selected cities. Men were included if they were born biologically male, aged 16 years and above, ever had sex with another man, and HIV negative or with unknown HIV status. All individuals completed informed consent. The intervention included individual- and community-level components. We sent campaign images and texts promoting HIV testing, along with local testing site information, to all participants privately on WeChat, an instant messaging tool (individual-level component). Local community-based organizations organized a crowdsourcing contest soliciting individual stories relevant to HIV testing among the MSM community with an aim of improving community engagement (community-level component). We followed the participants quarterly for 1 year. We conducted the final follow-up survey immediately after the completion of the final round of intervention. We conducted a secondary analysis of the final follow-up survey data for this study.

Measures

Sociodemographic, Sexual History, and Intervention Exposure Measurement

We collected information about men's sociodemographic characteristics: age, residence status (nonmigrants or migrants), marital status (never married, currently married, divorced or widowed), educational attainment (high school or below, some college, and college or above), and annual income (US $<2500, US $2501-8500, US $8501-14,000, and US $>14,000). Sexual history included their sexual orientation (gay, bisexual, or unsure), sexual orientation disclosure to others (yes or no), sexual orientation disclosure to health care providers (yes or no), number of male sex partners in the last 3 months, whether had regular and/or casual male partners in the last 3 months (yes or no), and whether met male sex partner online, including website and social media platforms (yes or no). Exposure to interventional materials, including images, texts, local testing sites information, and a local crowdsourcing contest promoting HIV testing, was also noted (yes or no).

Web-Based Sexual Health Influence Measurement

Personal influence, or communication between the communicator and receiver, has been noted historically as a powerful factor in explaining and predicting people's behavior [31]. It has been conceptualized as opinion leadership in studies of the diffusion process, while this diffusion leads to people's adoption of a new idea, behavior, or product [14]. We adapted the scale from a standardized 6-item opinion leadership scale [31] to measure *Web-based SHI* (abbreviated in text as *influencers*). A similar 6-item scale had been used to assess opinion leadership among a Web-based Taiwanese MSM population [19]. The scale assessed men's influence on spreading HIV and STI information online. The 6 items included: (1) how often they talked to their MSM friends or followers about HIV and STI, (2) how much information they provided when talking about HIV and STI, (3) how many MSM they told about HIV and STI, (4) how likely they were to be asked for more information about HIV and STI, (5) who communicated more information about HIV and STI: the participants or their MSM friends or followers, or almost equal, and (6) how often they were used as a source of advice. All items were assessed on a 5-point Likert scale. The Cronbach alpha of the scale is .937 in this study. On the basis of Rogers' diffusion of innovation theory, approximately 15% of the individuals of a community are early adopters of an innovation and can consequently influence others as well as shape social norms [14]. We defined *Web-based SHIs* as those whose opinion leadership mean score ranked within the top 15% (mean score >3) in the cohort (mean score ranged from 1 to 5). However, because of difficulty with multiple observations showing the same score at the 15% cut-off, a slightly tighter but most close cut-off 13% was chosen, rather than 15%.

Behavioral and Social Media Engagement Measurement

Men were asked about any condomless sex with male partners in the last 3 months (yes or no), HIV testing in the last 3 months including either facility-based testing or self-testing (yes or no), and syphilis testing in the last 3 months (yes or no). We asked about their social media engagement, defined as whether they reported using Weibo (microblog similar to Twitter), WeChat, and QQ (both are instant messaging mobile apps) or Blued (a gay dating app) in the last 3 months to give or receive information about HIV testing, except for the information delivered by the trial (yes or no). We asked men to report the number of followers on their various social media platforms, including Weibo, WeChat, QQ, and Blued. We also asked the respondents to self-report approximately how many of their MSM friends or followers on social media have gone for an HIV test after their intervention, using a 5-point Likert scale question (none at all, a few, some, many, or quite a lot).

Community Engagement and HIV-Relevant Psychological Measurement

We measured community engagement, anticipated HIV stigma [32], HIV testing social norms [33], and HIV testing self-efficacy [34]. Community engagement was measured by 6 items, with each using binary responses (yes or no), which assessed men's level of engagement in sexual health activities. Score of community engagement ranged from 0 to 6, and a higher score means better community engagement in sexual health. The 7-item anticipated HIV stigma scale asked participants about their own feelings about themselves if they had HIV as well as perceived discriminating attitudes from other people. HIV testing social norms were measured using a validated 6-item scale asking about men's opinions of the gay community's attitudes toward HIV testing. HIV testing self-efficacy was measured using a 6-item scale. Answers to items of these scales were given in a 4-point Likert format: strongly disagree=1, disagree=2, agree=3, and strongly agree=4. Mean scores for anticipated HIV stigma, HIV testing social norms, and self-efficacy were calculated by averaging the summed values of all items, ranging from 1 to 4. A higher score means a higher level of anticipated stigma, better perceived

social norms, or better self-efficacy. All scales have been used and evaluated in China before. In this study, the Cronbach alpha values for community engagement, anticipated HIV stigma, HIV testing social norms, and HIV testing self-efficacy were .709, .880, .494, and .787, respectively (see Multimedia Appendix 1 for more detailed responses to individual items of the scales).

Statistical Analysis

Due to the small cells of self-reported influence on others' adoption of HIV test within influencers, we dichotomized the variable and grouped *many* and *quite a lot* into 1 category for further analysis. Descriptive analysis was conducted to describe the differences in sociodemographic characteristics, sexual behaviors, HIV and syphilis testing history, exposure to interventional materials, use of social media platforms, anticipated HIV-related stigma, HIV testing social norms, HIV testing self-efficacy, and community engagement between influencers and noninfluencers. Chi-square or Fisher exact test was used for categorical variables, and independent sample *t* test was used for continuous variables.

We conducted multivariable linear or logistic regression analyses to examine the association between opinion leadership and various characteristics outcomes aforementioned. Mean differences in continuous variables between influencers and noninfluencers were evaluated, whereas odds ratios were reported for binary outcomes. Given the effects of trial interventions on these outcomes as well as its potential association with leadership, we controlled all previous interventions as predictors in the crude model (model 1). In addition, sociodemographic factors including age, education, income, and marital status were controlled in an extended model (model 2). We reported 95% CIs and *P* values. A *P* value of less than .05 was considered statistically significant. Data were analyzed with SAS software, version 9.4 (SAS Institute, Cary, USA).

Ethical Statement

Ethical approval was obtained from the institutional review committees at the Dermatology Hospital of Southern Medical University (Guangzhou, China), Shandong University (Jinan, China), University of North Carolina at Chapel Hill (Chapel Hill, North Carolina), and the University of California, San Francisco (San Francisco, California) before the launch of the survey.

Results

Sociodemographic and Behavioral Characteristics

Among 1031 men, 132 (132/1031, 12.80%) had a mean leadership score of greater than 3 and were categorized as influencers and 899 (899/1031, 87.20%) were noninfluencers (Multimedia Appendix 1). Most men were younger than 30 years (819/1031, 79.44%), were living in the sampling city (926/1031, 89.82%), were never married (907/1031, 87.97%), obtained college or above education (667/1031, 64.69%), and self-identified as gay (741/1031, 71.87%; see Table 1). Compared with noninfluencers, influencers were more likely to disclose their sexual orientation to others (100/132, 75.8% vs 580/899, 64.5%; *P*=.01), disclose sexual orientation to health care providers (37/132, 28.0% vs 178/899, 19.8%; *P*=.03), have a casual male partner in the last 3 months (63/132, 47.7% vs 315/899, 35.0%; *P*=.05), test for HIV in the last 3 months (73/132, 55.3% vs 337/899, 37.5%; *P*<.001), test for syphilis in the last 3 months (35/132, 26.5% vs 137/899, 15.2%; *P*=.001), and self-report that many or quite a lot of others had taken an HIV test after their intervention within their network (42/132, 31.8% vs 38/899, 4.2%; *P*<.001). There were no significant differences between the 2 groups in terms of the average number of male sex partners (1.32 vs 1.11; *P*=.16), condomless sex with male partners in the last 3 months (26/132, 19.7% vs 203/899, 22.6%; *P*=.46), and mainly meeting male sex partners online in the last 12 months (97/132, 73.5% vs 669/899, 74.4%; *P*=.82).

Social Media Engagement and Participation in Program Intervention

Table 2 shows exposure to the intervention trial and social media engagement by influencers and noninfluencers. Being in the same cohort, influencers were more likely to have seen our interventional materials, including images promoting HIV testing (125/132, 94.7% vs 759/899, 84.4%; *P*=.002), texts promoting HIV testing (120/132, 90.9% vs 672/899, 74.7%; *P*<.001), and a local crowdsourcing contest (63/132, 47.7% vs 312/899, 34.7%; *P*=.004). In terms of using social media platforms to give and receive any extra information relevant to HIV testing (except for the information delivered by the trial), influencers were more likely to give or receive extra information than noninfluencers on Weibo (42/132, 31.8% vs 112/899, 12.5%; *P*<.001), WeChat (85/132, 64.4% vs 275/899, 30.6%; *P*<.001), QQ (54/132, 40.9% vs 125/899, 13.9%; *P*<.001%), and mobile apps (42/132, 31.8% vs 169/899, 18.8% *P*=.001). As shown in Table 3, compared with noninfluencers, influencers had significantly more QQ followers (238 vs 159; *P*=.03), less Blued followers (172 vs 466; *P*=.003), a lower anticipated HIV stigma score (2.7 vs 2.9; *P*<.001), a higher HIV testing self-efficacy score (3.4 vs 3.1; *P*<.001), and a higher community engagement score (4.0 vs 2.5; *P*<.001).

Factors Related to Being a Sexual Health Influencer

As shown in Table 4, being influencers was associated with lower mean scores of anticipated HIV stigma (mean difference −0.22; *P*=.02), higher mean scores of HIV testing self-efficacy (mean difference 0.24; *P*=.01), and higher summed scores of community engagement (mean difference 1.48; *P*<.001). The differences remained significant after controlling for sociodemographic factors additionally in model 2.

Table 1. Demographic characteristics of Web-based sexual health influencers and noninfluencers in a men who have sex with men cohort in China in 2016 to 2017.

Demographic characteristics	Total (N=1031)	Noninfluencers (N=899)	Influencers (N=132)	*P* value
Age (years), n (%)				.41
<20	174 (16.88)	154 (17.1)	20 (15.2)	
20-29	645 (62.56)	555 (61.7)	90 (68.2)	
30-39	170 (16.49)	154 (17.1)	16 (12.1)	
≥40	42 (4.07)	36 (4.0)	6 (4.5)	
Residence status , n (%)				.27
Nonmigrants	926 (89.82)	811 (90.2)	115 (87.1)	
Migrants	105 (10.18)	88 (9.8)	17 (12.9)	
Marital status, n (%)				.96
Never married	907 (87.97)	790 (87.9)	117 (88.6)	
Currently married	89 (8.63)	78 (8.7)	11 (8.3)	
Divorced or widowed	35 (3.40)	31 (3.4)	4 (3.0)	
Educational level attained, n (%)				.17
High school or below	364 (35.31)	327 (36.4)	37 (28.0)	
Some college	285 (27.64)	245 (27.3)	40 (30.3)	
College or above	382 (37.05)	327 (36.4)	55 (41.7)	
Annual income (US $), n (%)				.17
≤2500	235 (22.79)	202 (22.5)	33 (25)	
2501-8500	544 (52.76)	474 (52.7)	70 (53.0)	
8501-14,000	159 (15.42)	146 (16.2)	13 (9.8)	
>14,000	93 (9.02)	77 (8.6)	16 (12.1)	
Sexual orientation, n (%)				.86
Homosexual	741 (71.87)	647 (72.0)	94 (71.2)	
Bisexual	252 (24.44)	218 (24.2)	34 (25.8)	
Unsure	38 (3.69)	34 (3.8)	4 (3.0)	
Disclose sexual orientation to anyone[a], n (%)	680 (65.96)	580 (64.5)	100 (75.8)	.01
Disclose sexual orientation to health providers[a], n (%)	215 (20.85)	178 (19.8)	37 (28.0)	.03
Number of male sex partners in the last 3 months, mean (SD)	1.14 (1.5)	1.11 (1.5)	1.32 (1.8)	.16
Mainly met male sex partners online in the last 12 months[a,b], n (%)	766 (74.30)	669 (74.4)	97 (73.5)	.82
Had a regular male partner in the last 3 months[a,c], n (%)	351 (34.04)	307 (34.1)	44 (33.3)	.85
Had a casual male partner in the last 3 months[a,d], n (%)	378 (36.66)	315 (35.0)	63 (47.7)	.005
Condomless sex with male partners in the last 3 months[a], n (%)	229 (22.21)	203 (22.6)	26 (19.7)	.46
HIV test in the last 3 months[a,e], n (%)	410 (39.77)	337 (37.5)	73 (55.3)	<.001
Syphilis test in the last 3 months[a], n (%)	172 (16.68)	137 (15.2)	35 (26.5)	.001
Self-reported influence on others' adoption of an HIV test after their Web-based intervention (Many or quite a lot), n (%)	80 (7.76)	38 (4.2)	42 (31.8)	<.001

[a]The response was yes for these variables.

[b]Mainly met with male sexual partners through a website or social media platforms.

[c]Regular male partner was defined as the one who was in a stable relationship (over 3 months) that did not involve transactional sex.

[d]Casual male partner was defined as a male sexual partner that the participant did not consider to be his regular partner.

[e]Either facility-based testing or self-testing.

Table 2. Exposure to the trial intervention reported by Web-based sexual health influencers and noninfluencers in a men who have sex with men cohort in China in 2016 to 2017.

Binary outcomes	Total (N=1031)	Noninfluencers (N=899)	Influencers (N=132)	*P* value
Exposure to the trial intervention materials, n (%)				
Have seen any images promoting HIV testing	884 (85.74)	759 (84.4)	125 (94.7)	.002
Have seen any texts promoting HIV testing	792 (76.82)	672 (74.7)	120 (90.9)	<.001
Have seen local testing sites information	864 (83.80)	749 (83.3)	115 (87.1)	.27
Have seen a local crowdsourcing contest	375 (36.37)	312 (34.7)	63 (47.7)	.004
Give or receive anything related to HIV testing on social media platforms except for the information from the trial, n (%)				
Using Weibo to give or receive information	154 (14.94)	112 (12.5)	42 (31.8)	<.001
Using WeChat to give or receive information	360 (34.92)	275 (30.6)	85 (64.4)	<.001
Using QQ to give or receive information	179 (17.36)	125 (13.9)	54 (40.9)	<.001
Using Blued to give or receive information	211 (20.47)	169 (18.8)	42 (31.8)	.001

Table 3. Number of social media followers, HIV-relevant psychological profiles, and community engagement by Web-based sexual health influencers and noninfluencers in a men who have sex with men cohort in China in 2016 to 2017 (N=1031).

Continuous outcomes	Noninfluencers	Influencers	*P* value
Number of Weibo followers, mean (SD)	269 (1430)	740 (5267)	.31
Number of WeChat followers, mean (SD)	168 (317)	749 (4725)	.17
Number of QQ followers, mean (SD)	159 (334)	238 (374)	.03
Number of Blued followers, mean (SD)	466 (2709)	172 (418)	.003
Anticipated HIV stigma[a], mean score (SD)	2.9 (0.7)	2.7 (0.8)	<.001
HIV testing social norms[b], mean score (SD)	2.9 (0.4)	2.8 (0.4)	.76
HIV testing self-efficacy[b], mean score (SD)	3.1 (0.5)	3.4 (0.5)	<.001
Community engagement[c], mean score (SD)	2.5 (1.7)	4.0 (1.6)	<.001

[a]Mean scores of anticipated HIV stigma ranged from 1 to 4, and a higher score means a higher level of anticipated stigma.

[b]Mean scores of HIV testing social norms and self-efficacy ranged from 1 to 4, and higher mean scores mean better perceived social norms and better self-efficacy.

[c]Score of community engagement ranged from 0 to 6, and a higher score means better community engagement in sexual health.

Table 4. Association between Web-based sexual health influence and continuous outcomes in a men who have sex with men cohort in China in 2016 to 2017 (N=1031).

Continuous outcomes[a]	Model 1[b]		Model 2[c]	
	Estimated mean difference (95% CI)	*P* value	Estimated mean difference (95% CI)	*P* value
Anticipated HIV stigma	−0.22 (−0.40 to −0.05)	.02	−0.23 (−0.35 to −0.12)	<.001
HIV testing social norms	−0.01 (−0.11 to 0.08)	.77	−0.02 (−0.10 to 0.06)	.69
HIV testing self-efficacy	0.24 (0.07 to 0.41)	.01	0.25 (0.16 to 0.34)	<.001
Community engagement	1.48 (1.06 to 1.90)	<.001	1.50 (1.19 to 1.81)	<.001

[a]Reference group is nonsexual health influencers.

[b]Model 1 was only adjusted for a previous intervention package to promote HIV testing among the cohort.

[c]Model 2 was additionally adjusted for age, education, income, and marital status.

Table 5. Association between Web-based sexual health influence and binary outcomes in a men who have sex with men cohort in China in 2016 to 2017 (N=1031).

Behavioral outcomes[a]	Model 1[b]		Model 2[c]	
	Estimated odds ratio (95% CI)	*P* value	Estimated odds ratio (95% CI)	*P* value
Testing behaviors and condom use in the last 3 months				
Overall HIV testing	2.12 (1.45-3.09)	<.001	2.16 (1.48-3.17)	<.001
HIV self-testing	1.62 (1.09-2.42)	.02	1.64 (1.10-2.46)	.02
HIV facility-based testing	2.59 (1.75-3.82)	<.001	2.66 (1.79-3.96)	<.001
Consistent condom use	1.29 (0.78-2.12)	.32	1.33 (0.80-2.19)	.27
Syphilis testing	1.94 (1.26-3.01)	<.01	1.99 (1.28-3.10)	<.01
Social media engagement[d]				
Using Weibo to give or receive information	1.88 (1.20-2.97)	<.001	1.90 (1.19-3.02)	<.01
Using WeChat to give or receive information	3.56 (1.78-7.13)	<.001	3.79 (1.87-7.66)	<.001
Using QQ to give or receive information	2.76 (1.71-4.45)	<.001	2.91 (1.79-4.75)	<.001
Using an app to give or receive information	1.07 (0.68-1.69)	.76	1.04 (0.66-1.64)	.87
Self-reported influence on others' HIV test uptake				
Many or quite a lot of people took a test	6.81 (4.14-11.2)	<.001	7.62 (4.55-12.78)	<.001

[a]Reference group is nonsexual health influencers.

[b]Model 1 was only adjusted for a previous intervention package to promote HIV testing among the cohort.

[c]Model 2 was additionally adjusted for age, education, income, and marital status.

[d]Social media engagement was defined as whether they reported using Weibo, WeChat, QQ, or a mobile app in the last 3 months to give or receive information about HIV testing, except for the information delivered by the trial.

Table 5 shows comparisons of binary outcomes between the 2 groups. Being influencers was found to be significantly associated with higher odds of HIV testing in the last 3 months (adjusted odds ratio, AOR 2.16, 95% CI 1.48-3.17), HIV self-testing in the last 3 months (AOR 1.64, 95% CI 1.10-2.46), HIV facility-based testing in the last 3 months (AOR 2.66, 95% CI 1.79-3.96), and syphilis testing in the last 3 months (AOR 1.95, 95% CI 1.25-3.03). Influencers were 1.90 (95% CI 1.19-3.02), 3.79 (95% CI 1.87-7.66), and 2.91 (95% CI 1.79-4.75) times more likely to use Weibo, WeChat, and QQ, respectively to give or receive extra HIV testing–relevant information than noninfluencers. In terms of self-reported influence on others' HIV test uptake, influencers were 7.62 (95% CI 4.55-12.78) times more likely to report that many or quite a lot of people within their network have gone for an HIV test after their intervention. However, opinion leadership was found not to be associated with consistent condom use.

Discussion

Principal Findings

Influential individuals may help promote health behaviors within their networks; however, there is insufficient understanding of the characteristics of Web-based influential individuals who are increasingly important to health promotion in the internet era. By examining social media engagement and health behaviors of Web-based SHIs among Chinese MSM, this study extends the literature by illuminating the degree to which existing influential individuals may be useful agents in the Web-based virtual space and by informing the development of network-based social media interventions. Our study found that influencers had higher social media engagement for HIV testing, higher likelihood of HIV and syphilis testing, and did not have increased sexual risk behavior. This group could become key leaders within network-based social media MSM HIV interventions.

We found that influencers had a higher HIV testing rate than noninfluencers. The HIV testing rate of influencers was higher than that of Chinese MSM in another postintervention study [35]. After adjusting for the intervention, influencers in this study also had a higher likelihood of HIV testing. Their higher HIV testing may be related to increased community engagement in sexual health [36] and HIV testing–related social media use [37]. The higher rates of HIV testing among influencers may also be related to lower HIV stigma and higher testing self-efficacy, which are 2 important contributors to test uptake [32,38]. Influencers also had a higher rate of syphilis testing than noninfluencers in the last 3 months (26.5% vs 15.2%). This is consistent with previous studies that found community engagement in sexual health to be associated with increased syphilis testing [36]. This suggests that influencers could be helpful in promoting dual HIV and syphilis testing, given that these related infections often co-occur among MSM in China [39].

In terms of influence on others' adoption of HIV test, we found influencers were significantly associated with reporting that many people within their Web-based social network have taken an HIV test after their intervention. This may be explained by greater social media engagement and larger social network sizes

among influencers. Influencers had a greater exposure to our trial intervention materials, including seeing any images or texts promoting HIV testing and local testing sites information. They were also more active in using various social media platforms to communicate with others about HIV testing. Being more active in receiving and disseminating sexual health information indicated that natural influencers may be more central to information flow, hence facilitating healthy behavior spread within their social network.

Finally, there was no significant difference between influencers and noninfluencers with regard to sexual behaviors. Specifically, rates for condomless sex, having a regular male partner, meeting sex partners online, and the mean number of male sex partners were similar. These findings are relevant to the growing body of research on internet use and sexual risk behaviors among MSM. On the one hand, internet use and social media are believed to be an avenue for meeting MSM, who then engage in risky behaviors associated with transmission of HIV and other STIs [40,41]. However, our study revealed that influencers had less followers than noninfluencers in Blued, the most popular gay social networking app in China. This indicates that influencers may in fact use this app less often for finding dates online. Furthermore, social media use could also potentially decrease sexual risk behaviors as it allows MSM to discuss with others about sexual health and learn about HIV and STI prevention [42]. Influencers in this study had a higher degree of online usage and communication with other MSM; however, they do not have increased sexual risk behaviors.

Limitations

We noted some limitations of the study. First, the study tended to focus on describing the influencers' characteristics and behaviors. Yet, descriptive studies on influencers are valuable, given that health behaviors are known to spread from person to person in social networks [43,44]. Our study may also provide a mechanism (ie, centrality in information flow) explaining why naturally existent influencers may facilitate behavioral change within their own network. Second, we only evaluated opinion leadership in the final follow-up survey and were not able to examine the longitudinal effects of the intervention on the relationship between opinion leadership and testing behaviors. Instead, we controlled for the intervention in the models to eliminate the confounding effects of interventions on the outcomes. Third, we used self-reported data to measure leadership and their influence on others' HIV test uptake. More reliable measurement and more robust research are necessary to evaluate the effect of influencers within MSM social networks with regard to behavior change. Incorporating social network measurement (eg, eigenvector centrality) and personal influence (eg, opinion leadership) and measuring their effects on positive behavior spread are worth consideration. Finally, our cross-sectional design makes it difficult to determine causal relationships.

Implications

Our findings have implications for strengthening HIV and syphilis testing interventions. Web-based SHIs could be useful in a range of testing promotion models, particularly network-based social media approaches. Vulnerable populations such as MSM may distrust outside authorities but find advice from known influencers who carry credibility [45]. For example, influencers could serve as steering committee members of crowdsourcing contests that aim to promote testing [30,46]. Influencers could also serve as leaders in network-based social media interventions to allow health messages to reach more MSM [37]. We found a higher rate of social media engagement about HIV testing and a higher likelihood of having used social media for HIV testing in influencers compared with other MSM, suggesting influencers may be readily incorporated into social media HIV interventions [37]. Influencers additionally contribute by being well positioned in social networks to spread behavior change. Interventions that operate through existing influencers' social networks hold promise for reaching vulnerable communities, particularly when formal prevention infrastructure supports are limited [17,47].

Acknowledgments

This study was supported by the National Institutes of Health and the Bill & Melinda Gates Foundation.

Authors' Contributions

JDT and CW led the project. CL, RF, and WH collected and cleaned data. WT and HL conducted statistical analyses. DW and AL wrote the first draft of this manuscript. DW and TPZ revised the sections on introduction and discussion. Other authors provided constructive comments and edited the manuscript. All authors approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed responses to individual items of scales used in the study.

[PDF File (Adobe PDF File), 69KB - jmir_v21i1e10171_app1.pdf]

References

1. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi: 10.1016/S0140-6736(12)60821-6] [Medline: 22819660]
2. Feng Y, Wu Z, Detels R, Qin G, Liu L, Wang X, et al. HIV/STD prevalence among men who have sex with men in Chengdu, China and associated risk factors for HIV infection. J Acquir Immune Defic Syndr 2010 Feb;53(Suppl 1):S74-S80. [doi: 10.1097/QAI.0b013e3181c7dd16] [Medline: 20104114]
3. Ma X, Zhang Q, He X, Sun W, Yue H, Chen S, et al. Trends in prevalence of HIV, syphilis, hepatitis C, hepatitis B, and sexual risk behavior among men who have sex with men. Results of 3 consecutive respondent-driven sampling surveys in Beijing, 2004 through 2006. J Acquir Immune Defic Syndr 2007 Aug 15;45(5):581-587. [doi: 10.1097/QAI.0b013e31811eadbc] [Medline: 17577125]
4. United Nations Programme on HIV and AIDS. 2016. Global AIDS Update 2016 URL: http://www.unaids.org/sites/default/files/media_asset/global-AIDS-update-2016_en.pdf [accessed 2018-12-06] [WebCite Cache ID 74SmqVSHS]
5. He Q, Wu XQ, Han DL, Liang Z. HIV infection and risk behavior of men having sex with men in Chengdu City from 2004 to 2007. J Occup Health Damage 2008(4):222-224 [FREE Full text]
6. Ding XB, Feng LG, Xiao Y, Jin CR, Xu SM, Guo XJ, et al. Treatment efficacy of behavior intervention on male homosexual AIDS patients. J Trop Med 2010:322-326.
7. National Health and Family Planning Commission of the People's Republic of China. United Nations Programme on HIV and AIDS. 2015 May. 2015 China AIDS Response Progress Report URL: http://www.unaids.org/sites/default/files/country/documents/CHN_narrative_report_2015.pdf [accessed 2018-12-06] [WebCite Cache ID 74SoNSIuF]
8. Chow EP, Lau JT, Zhuang X, Zhang X, Wang Y, Zhang L. HIV prevalence trends, risky behaviours, and governmental and community responses to the epidemic among men who have sex with men in China. Biomed Res Int 2014;2014:607261 [FREE Full text] [doi: 10.1155/2014/607261] [Medline: 24822214]
9. Zou H, Hu N, Xin Q, Beck J. HIV testing among men who have sex with men in China: a systematic review and meta-analysis. AIDS Behav 2012 Oct;16(7):1717-1728. [doi: 10.1007/s10461-012-0225-y] [Medline: 22677975]
10. Best J, Tang W, Zhang Y, Han L, Liu F, Huang S, et al. Sexual behaviors and HIV/syphilis testing among transgender individuals in China: implications for expanding HIV testing services. Sex Transm Dis 2015 May;42(5):281-285 [FREE Full text] [doi: 10.1097/OLQ.0000000000000269] [Medline: 25868142]
11. Chow EP, Wilson DP, Zhang L. The rate of HIV testing is increasing among men who have sex with men in China. HIV Med 2012 May;13(5):255-263 [FREE Full text] [doi: 10.1111/j.1468-1293.2011.00974.x] [Medline: 22252151]
12. Tucker JD, Muessig KE, Cui R, Bien CH, Lo EJ, Lee R, et al. Organizational characteristics of HIV/syphilis testing services for men who have sex with men in South China: a social entrepreneurship analysis and implications for creating sustainable service models. BMC Infect Dis 2014 Nov 25;14:601 [FREE Full text] [doi: 10.1186/s12879-014-0601-5] [Medline: 25422065]
13. Valente T. Network models of the diffusion of innovations. Comput Math Organiz Theor 1996;2(2):163-164. [doi: 10.1007/BF00240425]
14. Rogers EM. Diffusion of Innovations. New York: Free Press; 2010:518.
15. Aral S, Walker D. Identifying influential and susceptible members of social networks. Science 2012 Jul 20;337(6092):337-341 [FREE Full text] [doi: 10.1126/science.1215842] [Medline: 22722253]
16. Aral S, Muchnik L, Sundararajan A. Engineering social contagions: optimal network seeding in the presence of homophily. Net Sci 2013 Jul 30;1(02):125-153. [doi: 10.1017/nws.2013.6]
17. Kim DA, Hwong AR, Stafford D, Hughes DA, O'Malley AJ, Fowler JH, et al. Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. Lancet 2015 Jul 11;386(9989):145-153. [doi: 10.1016/S0140-6736(15)60095-2] [Medline: 25952354]
18. Kelly JA. Popular opinion leaders and HIV prevention peer education: resolving discrepant findings, and implications for the development of effective community programmes. AIDS Care 2004 Feb;16(2):139-150. [doi: 10.1080/09540120410001640986] [Medline: 14676020]
19. Ko NY, Hsieh CH, Wang MC, Lee C, Chen CL, Chung AC, et al. Effects of internet popular opinion leaders (iPOL) among internet-using men who have sex with men. J Med Internet Res 2013 Feb 25;15(2):e40 [FREE Full text] [doi: 10.2196/jmir.2264] [Medline: 23439583]
20. Duke JC, Hansen H, Kim AE, Curry L, Allen J. The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study. J Med Internet Res 2014 Jul 10;16(7):e169 [FREE Full text] [doi: 10.2196/jmir.3430] [Medline: 25014311]
21. Young SD, Holloway I, Jaganath D, Rice E, Westmoreland D, Coates T. Project HOPE: online social network changes in an HIV prevention randomized controlled trial for African American and Latino men who have sex with men. Am J Public Health 2014 Sep;104(9):1707-1712. [doi: 10.2105/AJPH.2014.301992] [Medline: 25033137]
22. Chang T, Chopra V, Zhang C, Woolford SJ. The role of social media in online weight management: systematic review. J Med Internet Res 2013 Nov 28;15(11):e262 [FREE Full text] [doi: 10.2196/jmir.2852] [Medline: 24287455]

23. Rosenberger JG, Reece M, Novak DS, Mayer KH. The internet as a valuable tool for promoting a new framework for sexual health among gay men and other men who have sex with men. AIDS Behav 2011 Apr;15(Suppl 1):S88-S90. [doi: 10.1007/s10461-011-9897-y] [Medline: 21331800]
24. Magee JC, Bigelow L, Dehaan S, Mustanski BS. Sexual health information seeking online: a mixed-methods study among lesbian, gay, bisexual, and transgender young people. Health Educ Behav 2012 Jun;39(3):276-289. [doi: 10.1177/1090198111401384] [Medline: 21490310]
25. Wilkerson JM, Smolenski DJ, Horvath KJ, Danilenko GP, Simon Rosser BR. Online and offline sexual health-seeking patterns of HIV-negative men who have sex with men. AIDS Behav 2010 Dec;14(6):1362-1370 [FREE Full text] [doi: 10.1007/s10461-010-9794-9] [Medline: 20799060]
26. Ramallo J, Kidder T, Albritton T, Blick G, Pachankis J, Grandelski V, et al. Exploring social networking technologies as tools for HIV prevention for men who have sex with men. AIDS Educ Prev 2015 Aug;27(4):298-311 [FREE Full text] [doi: 10.1521/aeap.2015.27.4.298] [Medline: 26241381]
27. Muessig KE, Bien CH, Wei C, Lo EJ, Yang M, Tucker JD, et al. A mixed-methods study on the acceptability of using eHealth for HIV prevention and sexual health care among men who have sex with men in China. J Med Internet Res 2015 Apr 21;17(4):e100 [FREE Full text] [doi: 10.2196/jmir.3370] [Medline: 25900881]
28. Wu Z, Xu J, Liu E, Mao Y, Xiao Y, Sun X, National MSM Survey Group. HIV and syphilis prevalence among men who have sex with men: a cross-sectional survey of 61 cities in China. Clin Infect Dis 2013 Jul;57(2):298-309 [FREE Full text] [doi: 10.1093/cid/cit210] [Medline: 23580732]
29. Akl A. Voice of America. 2017 Mar 7. Youth at Forefront of Global Internet Use URL: https://blogs.voanews.com/techtonics/2014/03/07/youth-at-forefront-of-global-internet-use/ [accessed 2018-12-06] [WebCite Cache ID 74SqdMmnH]
30. SESH Study Group, Tucker JD. Crowdsourcing to promote HIV testing among MSM in China: study protocol for a stepped wedge randomized controlled trial. Trials 2017 Oct 2;18(1):447 [FREE Full text] [doi: 10.1186/s13063-017-2183-1] [Medline: 28969702]
31. Childers TL. Assessment of the psychometric properties of an opinion leadership scale. J Mark Res 1986 May;23(2):184-188. [doi: 10.2307/3151666]
32. Golub SA, Gamarel KE. The impact of anticipated HIV stigma on delays in HIV testing behaviors: findings from a community-based sample of men who have sex with men and transgender women in New York City. AIDS Patient Care STDS 2013 Nov;27(11):621-627 [FREE Full text] [doi: 10.1089/apc.2013.0245] [Medline: 24138486]
33. Pettifor A, MacPhail C, Suchindran S, Delany-Moretlwe S. Factors associated with HIV testing among public sector clinic attendees in Johannesburg, South Africa. AIDS Behav 2010 Aug;14(4):913-921. [doi: 10.1007/s10461-008-9462-5] [Medline: 18931903]
34. Gu J, Lau JT, Tsui H. Psychological factors in association with uptake of voluntary counselling and testing for HIV among men who have sex with men in Hong Kong. Public Health 2011 May;125(5):275-282. [doi: 10.1016/j.puhe.2011.01.010] [Medline: 21419466]
35. Tang W, Han L, Best J, Zhang Y, Mollan K, Kim J, et al. Crowdsourcing HIV test promotion videos: a noninferiority randomized controlled trial in China. Clin Infect Dis 2016 Dec 1;62(11):1436-1442 [FREE Full text] [doi: 10.1093/cid/ciw171] [Medline: 27129465]
36. Zhang TP, Liu C, Han L, Tang W, Mao J, Wong T, et al. Community engagement in sexual health and uptake of HIV testing and syphilis testing among MSM in China: a cross-sectional online survey. J Int AIDS Soc 2017 Dec 3;20(1):21372 [FREE Full text] [doi: 10.7448/IAS.20.01.21372] [Medline: 28406270]
37. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, et al. Social media interventions to promote HIV testing, linkage, adherence, and retention: systematic review and meta-analysis. J Med Internet Res 2017 Nov 24;19(11):e394 [FREE Full text] [doi: 10.2196/jmir.7997] [Medline: 29175811]
38. Berendes S, Rimal RN. Addressing the slow uptake of HIV testing in Malawi: the role of stigma, self-efficacy, and knowledge in the Malawi BRIDGE Project. J Assoc Nurses AIDS Care 2011;22(3):215-228. [doi: 10.1016/j.jana.2010.08.005] [Medline: 21185751]
39. Chen G, Cao Y, Yao Y, Li M, Tang W, Li J, et al. Syphilis incidence among men who have sex with men in China: results from a meta-analysis. Int J STD AIDS 2017 Dec;28(2):170-178 [FREE Full text] [doi: 10.1177/0956462416638224] [Medline: 26992411]
40. Young SD, Szekeres G, Coates T. The relationship between online social networking and sexual risk behaviors among men who have sex with men (MSM). PLoS One 2013;8(5):e62271 [FREE Full text] [doi: 10.1371/journal.pone.0062271] [Medline: 23658716]
41. McFarlane M, Bull SS, Rietmeijer CA. The internet as a newly emerging risk environment for sexually transmitted diseases. J Am Med Assoc 2000 Jul 26;284(4):443-446. [Medline: 10904506]
42. Moreno MA, Parks MR, Zimmerman FJ, Brito TE, Christakis DA. Display of health risk behaviors on MySpace by adolescents: prevalence and associations. Arch Pediatr Adolesc Med 2009 Jan;163(1):27-34. [doi: 10.1001/archpediatrics.2008.528] [Medline: 19124700]
43. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379. [doi: 10.1056/NEJMsa066082] [Medline: 17652652]

44. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008 May 22;358(21):2249-2258 [FREE Full text] [doi: 10.1056/NEJMsa0706154] [Medline: 18499567]
45. Kelly JA, Amirkhanian YA, Kabakchieva E, Vassileva S, Vassilev B, McAuliffe TL, et al. Prevention of HIV and sexually transmitted diseases in high risk social networks of young Roma (Gypsy) men in Bulgaria: randomised controlled trial. Br Med J 2006 Nov 25;333(7578):1098 [FREE Full text] [doi: 10.1136/bmj.38992.478299.55] [Medline: 17040924]
46. Zhang Y, Kim JA, Liu F, Tso LS, Tang W, Wei C, et al. Creative contributory contests to spur innovation in sexual health: 2 cases and a guide for implementation. Sex Transm Dis 2015 Nov;42(11):625-628 [FREE Full text] [doi: 10.1097/OLQ.0000000000000349] [Medline: 26462186]
47. Amirkhanian YA, Kelly JA, Takacs J, McAuliffe TL, Kuznetsova AV, Toth TP, et al. Effects of a social network HIV/STD prevention intervention for MSM in Russia and Hungary: a randomized controlled trial. AIDS 2015 Mar 13;29(5):583-593 [FREE Full text] [doi: 10.1097/QAD.0000000000000558] [Medline: 25565495]

Abbreviations

AOR: adjusted odds ratio
MSM: men who have sex with men
POL: popular opinion leader
SHI: sexual health influencer
STI: sexually transmitted infection

Edited by G Eysenbach; submitted 19.02.18; peer-reviewed by M Hogben, Y Rykov; comments to author 17.08.18; revised version received 24.10.18; accepted 26.10.18; published 21.01.19.

Please cite as:
Wu D, Tang W, Lu H, Zhang TP, Cao B, Ong JJ, Lee A, Liu C, Huang W, Fu R, Li K, Pan SW, Zhang Y, Fu H, Wei C, Tucker JD
Leading by Example: Web-Based Sexual Health Influencers Among Men Who Have Sex With Men Have Higher HIV and Syphilis Testing Rates in China
J Med Internet Res 2019;21(1):e10171
URL: http://www.jmir.org/2019/1/e10171/
doi:10.2196/10171
PMID:30664490

Short Paper

Assessing the Impact of a Social Marketing Campaign on Program Outcomes for Users of an Internet-Based Testing Service for Sexually Transmitted and Blood-Borne Infections: Observational Study

Mark Gilbert[1,2], MD; Travis Salway[1,2], PhD; Devon Haag[1], MSc; Michael Kwag[3], BA; Joshua Edward[4], PhD; Mark Bondyra[1]; Joseph Cox[5], MD; Trevor A Hart[6,7], PhD; Daniel Grace[7], PhD; Troy Grennan[1,8], MD; Gina Ogilvie[1,2], MD; Jean Shoveller[2], PhD

[1]British Columbia Centre for Disease Control, Vancouver, BC, Canada

[2]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[3]Community Based Research Centre for Gay Men's Health, Vancouver, BC, Canada

[4]Health Initiative for Men, Vancouver, BC, Canada

[5]Department of Epidemiology, Biostatistics and Occupational Health, Faculty of Medicine, McGill University, Montreal, QC, Canada

[6]Department of Psychology, Ryerson University, Toronto, ON, Canada

[7]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[8]Division of Infectious Diseases, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Mark Gilbert, MD
British Columbia Centre for Disease Control
655 West 12th Avenue
Vancouver, BC, V5Z 4R4
Canada
Phone: 1 604 707 5619
Fax: 1 604 707 2401
Email: mark.gilbert@bccdc.ca

Abstract

Background: While social marketing (SM) campaigns can be effective in increasing testing for sexually transmitted and blood-borne infections (STBBIs), they are seldom rigorously evaluated and often rely on process measures (eg, Web-based ad click-throughs). With Web-based campaigns for internet-based health services, there is a potential to connect campaign process measures to program outcomes, permitting the assessment of venue-specific yield based on health outcomes (eg, click-throughs per test).

Objective: This study aims to evaluate the impact of an SM campaign by the promotional venue on use and diagnostic test results of the internet-based STBBI testing service GetCheckedOnline.com (GCO).

Methods: Through GCO, clients create an account using an access code, complete a risk assessment, print a lab form, submit specimens at a lab, and get results online or by phone. From April to August 2015, a campaign promoted GCO to gay, bisexual, and other men who have sex with men in Vancouver, Canada. The campaign highlighted GCO's convenience in 3 types of promotional venues—location advertisements in print or video displayed in gay venues or events, ads on a queer news website, and ads on geosocial websites and apps. Where feasible, individuals were tracked from campaign exposures to account creation and testing using venue-specific GCO access codes. In addition, Web-based ads were linked to alternate versions of the campaign website, which used URLs with embedded access codes to connect ad exposure to account creation. Furthermore, we examined the number of individuals creating GCO accounts, number tested, and cost per account created and test for each venue type.

Results: Over 6 months, 177 people created a GCO account because of the campaign, where 22.0% (39/177) of these completed testing; the overall cost was Can $118 per account created and Can $533 per test. Ads on geosocial websites and apps accounted for 46.9% (83/177) of all accounts; ads on the news website had the lowest testing rate and highest cost per test. We observed

variation between different geosocial websites and apps with some ads having high click-through rates yet low GCO account creation rates, and vice versa.

Conclusions: Developing mechanisms to track individuals from Web-based exposure to SM campaigns to outcomes of internet-based health services permits greater evaluation of the yield and cost-effectiveness of different promotional efforts. Web-based ads with high click-through rates may not have a high conversion to service use, the ultimate outcome of SM campaigns.

(J Med Internet Res 2019;21(1):e11291) doi:10.2196/11291

KEYWORDS

advertisements; diagnostic tests; health promotion; internet; men who have sex with men; social marketing

Introduction

Social marketing (SM) campaigns promoting testing for sexually transmitted and blood-borne infections (STBBIs) can effectively increase the uptake of testing [1,2]. However, SM testing campaigns are rarely evaluated rigorously owing to pressures of real-world implementation (eg, evaluation budget and difficulty determining campaign-specific effects in an exposed population) [1,3]. Web-based elements of SM campaigns are often evaluated through monitoring the number of views (impressions) of Web-based ads and comparing the proportions of individuals clicking through to visit (click-through rate) and use a website or service (conversion rate) [4,5]; this information is used to identify promotional venues with higher yield, allowing redirection of efforts to optimize campaign reach and inform future campaigns [1].

More robust evaluations of SM campaigns are possible for campaigns promoting internet-based health services, where users are tracked through service progression. If designed appropriately, campaign evaluations can follow individuals from their initial campaign ad view through to their program outcomes, permitting an assessment of yield of different venues based on actual health outcomes. This paper aims to describe the results of using such a design to evaluate the impact of an SM campaign on increasing the uptake of GetCheckedOnline.com (GCO) [6], an internet-based STBBI testing service in British Columbia (BC), Canada.

Methods

GetCheckedOnline

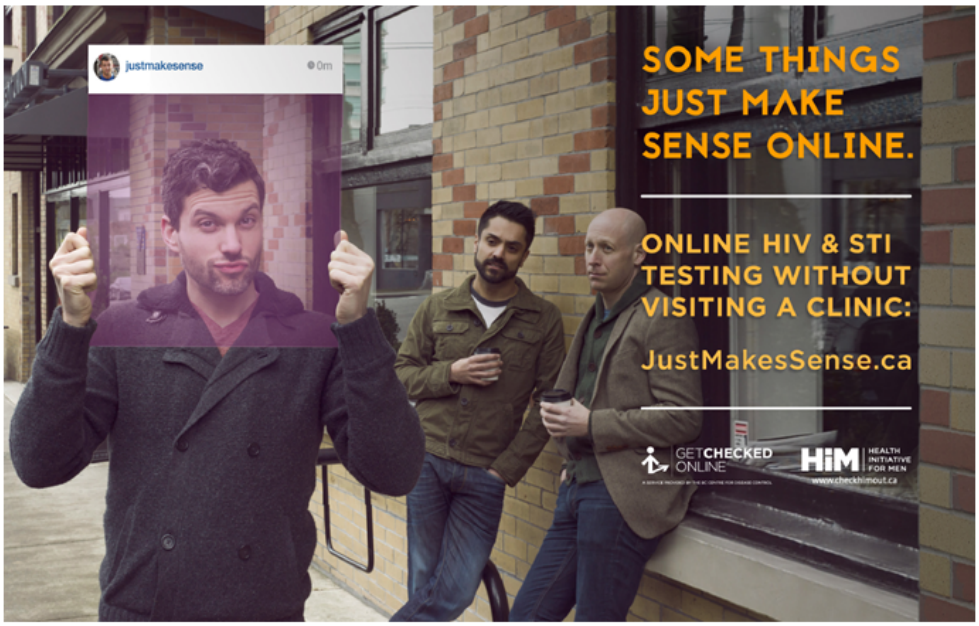
GCO is an internet-based testing service for STBBIs developed by the BC Centre for Disease Control (BCCDC), with a goal of overcoming existing testing barriers among populations with high rates of infection. We have previously published a full description of the GCO program [7]. In brief, users go through the following 5 steps to test through GCO: (1) create an account; (2) complete a risk assessment; (3) print a laboratory requisition form; (4) provide specimens in-person at a private lab (with testing for HIV, syphilis, hepatitis C, chlamydia, and gonorrhea); and (5) receive results online if negative, or by phone if positive or indeterminate. GCO accounts are created by entering an access code on the home page unique to a specific promotion strategy or venue. In addition, individuals can be invited to use GCO by emails with a link to the account creation page.

GCO was launched in 2014, initially targeting gay, bisexual, and other men who have sex with men (GBMSM) in the Vancouver region. Most GBMSM in BC regularly test, with 64.57% (1179/1826) and 62.28% (999/1604) reporting sexually transmitted infection and HIV testing, respectively, in the past year [8]. However, many GBMSM report delaying testing owing to barriers, including privacy concerns or inability to access clinics [9]. In formative research, GBMSM found GCO acceptable with high intention to use, particularly among men facing testing barriers, perceiving benefits, including greater privacy, convenience, control over testing, and not needing to see a health care provider [9,10].

Campaign Development

The SM campaign aimed to increase the awareness and uptake of GCO among GBMSM. The BCCDC partnered with the Health Initiative for Men, a community-based gay men's health organization, which led the development and implementation of the campaign in consultation with an advisory committee of GBMSM, sexual health nurses, and a small convenience sample of Health Initiative for Men clinic clients. The campaign focused on promoting the convenience of GCO, aiming to reach GBMSM avoiding or delaying testing because of access-related barriers (eg, wait-times for appointments). The campaign concept (Figure 1) was "Some things just make sense online," designed to use humor based on popular social media sites to motivate viewers to visit the JustMakesSense (JMS) campaign website [11] which emphasized the convenience and confidentiality of the service. Campaign materials included a website, videos, Web-based ads, and print media.

Figure 1. Examples of promotional campaign materials used for an internet-based testing service for sexually transmitted and blood-borne infections (GetCheckedOnline.com), including online banner advertisements (top), and in print (bottom).

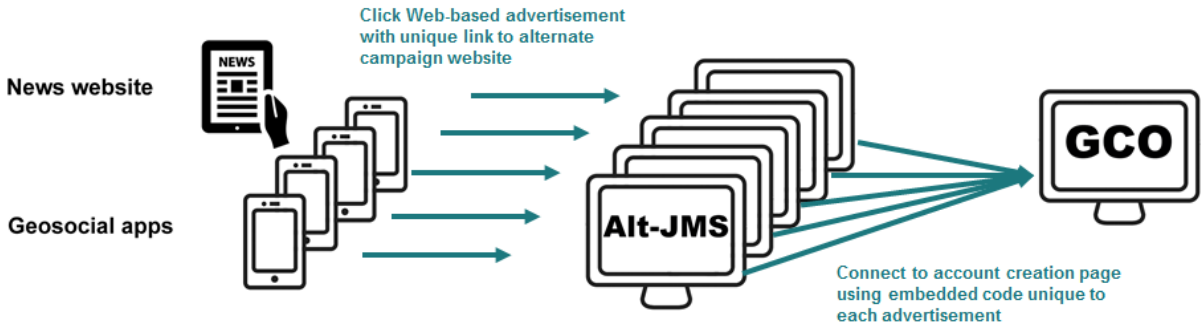


Tracking Program Outcomes by the Promotional Venue

The campaign ran from April to August 2015. We used 3 venue types for promotion, each having a unique route to account creation, permitting us to track testing outcomes (Figures 2 and 3). *Location ads* included the JMS website address and an access code unique to each location (ie, gay bars and clubs; sex on premises venues; community spaces; businesses; and a lesbian, gay, bisexual, and transgender film festival). Codes were short, easy-to-remember phrases, such as “TestNow” or “TestOnline,” displayed on videos, posters, or postcards (the latter could be taken home by individuals). On the JMS website, visitors entered an access code and proceeded to the account creation page on GCO. Visitors without a code could request an email invitation; visitors were not asked whether or where they had seen the campaign (Figure 2). In addition, we used 2 types of Web-based promotional venues—advertising in a *lesbian, gay, bisexual and transgender news website*, and advertising on *geosocial websites and apps* used by GBMSM to find sex partners (Grindr, Jack’d, Manhunt, Squirt, and Scruff). To track testing outcomes for each Web-based venue, each post or ad contained a link to a unique, alternate copy of the JMS website. From each alternate site, visitors proceeded to the GCO account creation page by clicking a link containing an embedded access code unique to each Web-based venue, which could then be associated with each account created.

Figure 2. The description of routes to account creation on GetCheckedOnline.com (GCO) during the JustMakesSense (JMS) campaign; visitors to the JMS main website.

Figure 3. The description of routes to account creation on GetCheckedOnline.com (GCO) during the JustMakesSense (JMS) campaign; visitors from online promotional venues to alternate JMS sites (Alt-JMS).



Data Analysis

Our primary outcome was the number of individuals creating GCO accounts by promotional venue type; secondary outcomes included the number of individuals tested, and costs per account created and individual tested. We collected available data from website or app vendors on impressions and click-through rates and extracted GCO program data. For each type of promotional venue, we calculated the number of GCO accounts created and proportion completing testing. For Web-based venues, we described the number of campaign impressions and click-throughs to alternate JMS campaign websites. Furthermore, we described the number of visitors creating accounts through requesting an invitation on the campaign website and their testing outcomes.

Ethics

Our analysis was conducted under a program evaluation mandate using data routinely collected by BCCDC or through contracts with Web-based ad vendors. The use of individual-level GCO program data is permitted for evaluation under the terms of service agreed to by all GCO users.

Results

Overall, 177 individuals created a GCO account because of the campaign, and 22.0% (39/177) of these completed testing; all results were negative (Table 1). The highest number of accounts was from individuals viewing campaign images on geosocial apps (83/177, 46.9%) followed by individuals requesting an invitation from the campaign website (52/177, 29.4%), location ads (21/177, 11.9%), and a news website (20/177, 11.3%). The completion of testing showed little variation across venues, except the news website (1/20, 5%). We spent Can $20,801 on promotion; the average cost was Can $118 per account created and Can $533 per test (Table 2). The costliest approach per account created was geosocial apps (Can $211), followed by Web-based news (Can $105) and location ads (Can $53). Web-based news had the highest cost per test (Can $2104). Over 19 million impressions of the JMS campaign occurred through geosocial apps, with the highest click-through rate on Grindr (0.7%). The highest numbers of accounts were created from ads on Manhunt and Squirt, resulting in low costs per account created (Can $83 and Can $213, respectively).

Table 1. Outcomes by the promotional venue.

Promotional venue	Number exposed to campaign or Number of impressions	Number visited JMS websites (% of click-through)	Number created account (% of visited Just-MakesSense site)	Number completed testing (% of accounts created)
Requested invitation	N/A[a]	N/A	52[b]	14 (26.92)
Location ads	N/A	N/A	21[c]	5 (23.80)
News website	195,120	260 (0.13)	20 (7.69)	1 (5.00)
Geosocial apps (all)	19,232,363	41,227 (0.21)	83 (0.20)	19 (22.89)
Grindr	3,443,423	24,975 (0.72)	5 (0.02)	2 (40.00)
Jack'd	366,744	655 (0.18)	1 (0.15)	0 (0.00)
Manhunt	299,284	547 (0.18)	33 (6.03)	11 (33.33)
Squirt	1,355,044	1,822 (0.13)	37 (2.03)	6 (16.21)
Scruff	13,767,868	13,228 (0.10)	7 (0.05)	0 (0.00)
Total	19,427,483	41,487 (0.21)	177 (0.30)	39 (22.03)

[a]N/A: not applicable.

[b]These individuals were not asked whether they were exposed to the campaign when they requested an invitation; hence we were unable to estimate the denominator (ie, the number exposed to the campaign) and report % values.

[c]We were unable to estimate the denominator (ie, the number exposed to location ads) and report % values for this.

Table 2. Costs by the promotional venue.

Promotional venue	Cost of promotion (Can $)	Cost per account (Can $)	Cost per test (Can $)
Location ads	1115	53	223
Web-based news	2104	105	2104
Geosocial apps (all)	17,582	211	925
Grindr	4270	854	2135
Jack'd	3200	3200	N/A[a]
Manhunt	7040	213	640
Squirt	3072	83	512
Scruff	0	0	N/A
Total	20,801	118	533

[a]N/A: not applicable.

Discussion

While commonly applied in e-commerce (eg, linking ad exposures to Web-based purchases), this study demonstrated the value of using this evaluation method to understand the effects of campaign ads (in a range of promotional venues) on internet-based health services. For example, just under half of all GCO accounts were created as a result of ads on geosocial apps, where we spent the bulk of our promotional budget. While Grindr had the highest click-through rate and would, therefore, typically be considered a successful promotional venue, Grindr had the lowest proportion of GCO accounts created and a higher cost per account and per test. Conversely, Manhunt and Squirt had lower click-through rates but had higher yield in terms of GCO program outcomes. In addition, we observed the highest account creation rate among individuals exposed to our Web-based news advertisement, although a much lower proportion proceeded to test. These differences in outcomes might be explained by several factors, including the characteristics of GBMSM on these different websites and apps, such as differences in response to the JMS campaign (influencing click-through rates); demographic factors (eg, age and ethnicity); and behavioral risk or testing barriers (influencing account creation and testing rates)—all aspects worthy of further study [12,13]. Furthermore, our findings demonstrate that the promotion in physical venues is important and cost-effective, as location ads had the lowest cost per account of all venues. However, we were unable to account for view through conversion, where GBMSM seeing campaign ads may have later requested a GCO invitation on the campaign website (29% of all accounts created).

We did not observe a large uptake in testing as a result of the JMS campaign. The 39 individuals testing through GCO may be "early adopters" of this intervention with the ongoing diffusion of this innovation through GBMSM networks [14]. A shift in the message may also be needed. The feedback from

GBMSM and providers following the campaign suggested convenience may not be the best selling point, given the relative availability of STBBI testing services for GBMSM in the Vancouver area (Edwards J, personal communication, November 2016); this may explain why only 1 in 5 men creating accounts tested through GCO, a measure associated with motivation to get tested in our prior evaluations [15].

In conclusion, this study demonstrates the value of developing mechanisms for tracking individuals from their Web-based exposure to SM campaign ads about an internet-based health service to their program outcomes. In addition, this study reveals that Web-based venues with high click-through rates may not always have a high conversion to service use, which is ultimately the desired outcome of SM campaigns. We are continuing to use venue-specific access codes to evaluate promotional efforts as GCO expands to other communities across BC.

Acknowledgments

Funding for the campaign and this evaluation was provided by the Canadian Institutes of Health Research (PHE-318068).

Authors' Contributions

MG, TS, DH, and MB conceived the overall evaluation design and worked with MK and JE on design, implementation, and data collection related to the JustMakeSense campaign. TS led the data analysis, and all authors provided input on the data analysis and interpretation. MG and TS led the drafting and revision of the manuscript, with input from all authors.

Conflicts of Interest

None declared.

References

1. Friedman AL, Kachur RE, Noar SM, McFarlane M. Health Communication and Social Marketing Campaigns for Sexually Transmitted Disease Prevention and Control: What Is the Evidence of their Effectiveness? Sex Transm Dis 2016 Feb;43(2 Suppl 1):S83-101. [doi: 10.1097/OLQ.0000000000000286] [Medline: 26779691]
2. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, et al. Social Media Interventions to Promote HIV Testing, Linkage, Adherence, and Retention: Systematic Review and Meta-Analysis. J Med Internet Res 2017 Nov 24;19(11):e394 [FREE Full text] [doi: 10.2196/jmir.7997] [Medline: 29175811]
3. Wei C, Herrick A, Raymond HF, Anglemyer A, Gerbase A, Noar SM. Social marketing interventions to increase HIV/STI testing uptake among men who have sex with men and male-to-female transgender women. Cochrane Database Syst Rev 2011(9):CD009337. [doi: 10.1002/14651858.CD009337] [Medline: 21901734]
4. Reiter PL, Katz ML, Bauermeister JA, Shoben AB, Paskett ED, McRee AL. Recruiting Young Gay and Bisexual Men for a Human Papillomavirus Vaccination Intervention Through Social Media: The Effects of Advertisement Content. JMIR Public Health Surveill 2017 Jun 02;3(2):e33 [FREE Full text] [doi: 10.2196/publichealth.7545] [Medline: 28576758]
5. Sanders A, Robinson C, Taylor SC, Post SD, Goldfarb J, Shi R, et al. Using a Media Campaign to Increase Engagement With a Mobile-Based Youth Smoking Cessation Program. Am J Health Promot 2018 Jun;32(5):1273-1279. [doi: 10.1177/0890117117728608] [Medline: 28925292]
6. GetCheckedOnline. URL: https://getcheckedonline.com/Pages/default.aspx [accessed 2018-06-12] [WebCite Cache ID 706UAkZuD]
7. Gilbert M, Haag D, Hottes TS, Bondyra M, Elliot E, Chabot C, et al. Get Checked… Where? The Development of a Comprehensive, Integrated Internet-Based Testing Program for Sexually Transmitted and Blood-Borne Infections in British Columbia, Canada. JMIR Res Protoc 2016 Sep 20;5(3):e186 [FREE Full text] [doi: 10.2196/resprot.6293] [Medline: 27649716]
8. Trussler T, Ham D. Community-based Research Centre for Gay Men's Health. Sex Now 2015 Health Authorities: Selected results from the National Sex Now Survey comparing BC Health Authorities URL: http://cbrc.net/sites/cbrc.net/files/sn_healthauth_en_170727_RGB_0.pdf [WebCite Cache ID 707kSwhs2]
9. Gilbert M, Hottes TS, Kerr T, Taylor D, Fairley CK, Lester R, et al. Factors associated with intention to use internet-based testing for sexually transmitted infections among men who have sex with men. J Med Internet Res 2013;15(11):e254 [FREE Full text] [doi: 10.2196/jmir.2888] [Medline: 24240644]
10. Hottes TS, Farrell J, Bondyra M, Haag D, Shoveller J, Gilbert M. Internet-based HIV and sexually transmitted infection testing in British Columbia, Canada: opinions and expectations of prospective clients. J Med Internet Res 2012;14(2):e41 [FREE Full text] [doi: 10.2196/jmir.1948] [Medline: 22394997]
11. Health Initiative for Men. Just Makes Sense campaign website URL: http://justmakessense.ca/ [accessed 2017-12-27] [WebCite Cache ID 6w1wBT2g3]
12. Sun CJ, Sutfin E, Bachmann LH, Stowers J, Rhodes SD. Comparing men who have sex with men and transgender women who use Grindr, other similar social and sexual networking apps, or no social and sexual networking apps: Implications

for recruitment and health promotion. J AIDS Clin Res 2018;9(2):757 [FREE Full text] [doi: 10.4172/2155-6113.1000757] [Medline: 29593933]
13. Badal HJ, Stryker JE, DeLuca N, Purcell DW. Swipe Right: Dating Website and App Use Among Men Who Have Sex With Men. AIDS Behav 2018 Apr;22(4):1265-1272. [doi: 10.1007/s10461-017-1882-7] [Medline: 28884248]
14. Haider M, Kreps GL. Forty years of diffusion of innovations: utility and value in public health. J Health Commun 2004;9 Suppl 1:3-11. [doi: 10.1080/10810730490271430] [Medline: 14960400]
15. Gilbert M, Salway T, Haag D, Fairley CK, Wong J, Grennan T, et al. Use of GetCheckedOnline, a Comprehensive Web-based Testing Service for Sexually Transmitted and Blood-Borne Infections. J Med Internet Res 2017 Dec 20;19(3):e81 [FREE Full text] [doi: 10.2196/jmir.7097] [Medline: 28320690]

Abbreviations

BC: British Columbia
BCCDC: British Columbia Centre for Disease Control
GBMSM: gay, bisexual and other men who have sex with men
GCO: GetCheckedOnline.com
JMS: JustMakeSense
SM: Social marketing
STBBI: sexually transmitted and blood-borne infection

Edited by G Eysenbach; submitted 26.06.18; peer-reviewed by KA Nguyen, SM Goodreau; comments to author 29.08.18; revised version received 12.09.18; accepted 14.09.18; published 21.01.19.

Please cite as:
Gilbert M, Salway T, Haag D, Kwag M, Edward J, Bondyra M, Cox J, Hart TA, Grace D, Grennan T, Ogilvie G, Shoveller J
Assessing the Impact of a Social Marketing Campaign on Program Outcomes for Users of an Internet-Based Testing Service for Sexually Transmitted and Blood-Borne Infections: Observational Study
J Med Internet Res 2019;21(1):e11291
URL: http://www.jmir.org/2019/1/e11291/
doi:10.2196/11291
PMID:30664456

Original Paper

Digital Recruitment and Acceptance of a Stepwise Model to Prevent Chronic Disease in the Danish Primary Care Sector: Cross-Sectional Study

Lars Bruun Larsen[1], MPH; Jens Sondergaard[1], PhD; Janus Laust Thomsen[2], PhD; Anders Halling[3], PhD; Anders Larrabee Sønderlund[1], PhD; Jeanette Reffstrup Christensen[1], PhD; Trine Thilsing[1], PhD

[1]Research Unit of General Practice, Institute of Public Health, University of Southern Denmark, Odense, Denmark

[2]Research Unit for General Practice, Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

[3]Department of Clinical Sciences in Malmö, Centre for Primary Health Care Research, Lund University, Lund, Sweden

Corresponding Author:
Lars Bruun Larsen, MPH
Research Unit of General Practice
Institute of Public Health
University of Southern Denmark
JB Winsløws Vej 9A
Odense, 5000
Denmark
Phone: 45 65503830
Email: lblarsen@health.sdu.dk

Abstract

Background: During recent years, stepwise approaches to health checks have been advanced as an alternative to general health checks. In 2013, we set up the Early Detection and Prevention project (Tidlig Opsporing og Forebyggelse, TOF) to develop a stepwise approach aimed at patients at high or moderate risk of a chronic disease. A novel feature was the use of a personal digital mailbox for recruiting participants. A personal digital mailbox is a secure digital mailbox provided by the Danish public authorities. Apart from being both safe and secure, it is a low-cost, quick, and easy way to reach Danish residents.

Objective: In this study we analyze the association between the rates of acceptance of 2 digital invitations sent to a personal digital mailbox and the sociodemographic determinants, medical treatment, and health care usage in a stepwise primary care model for the prevention of chronic diseases.

Methods: We conducted a cross-sectional analysis of the rates of acceptance of 2 digital invitations sent to randomly selected residents born between 1957 and 1986 and residing in 2 Danish municipalities. The outcome was acceptance of the 2 digital invitations. Statistical associations were determined by Poisson regression. Data-driven chi-square automatic interaction detection method was used to generate a decision tree analysis, predicting acceptance of the digital invitations.

Results: A total of 8814 patients received an invitation in their digital mailbox from 47 general practitioners. A total of 40.22% (3545/8814) accepted the first digital invitation, and 30.19 % (2661/8814) accepted both digital invitations. The rates of acceptance of both digital invitations were higher among women, older patients, patients of higher socioeconomic status, and patients not diagnosed with or being treated for diabetes mellitus, chronic obstructive pulmonary disease, or cardiovascular disease.

Conclusions: To our knowledge, this is the first study to report on the rates of acceptance of digital invitations to participate in a stepwise model for prevention of chronic diseases. More studies of digital invitations are needed to determine if the acceptance rates seen in this study should be expected from future studies as well. Similarly, more research is needed to determine whether a multimodal recruitment approach, including digital invitations to personal digital mailboxes will reach hard-to-reach subpopulations more effectively than digital invitations only.

(J Med Internet Res 2019;21(1):e11658) doi:10.2196/11658

KEYWORDS

promotion of health; clinical decision support systems; cross-sectional studies

Introduction

Background

General health checks are seen as one way to mitigate the rising prevalence of chronic diseases such as cardiovascular disease (CVD), type 2 diabetes mellitus (T2DM), and chronic obstructive pulmonary disease (COPD). Consequently, periodical general health checks are provided to citizens by various national health care systems around the world, including those in the United States, South Korea, Australia, and Germany. However, general health checks have not only failed to show population health effects on CVD and total mortality but may have also widen health inequalities [1-6]. This is probably because of the generally higher uptake of health check initiatives in the populations who are likely to benefit the least—including most notably women and patients of higher age, better health, and higher socioeconomic status (SES). Population-level uptake seems to be determined by an interrelationship between individual and societal facilitators and barriers as well as self-selection [7,8].

During recent years, stepwise approaches to health checks have been advanced as an alternative to general health checks [9,10]. Stepwise approaches usually comprise a risk assessment to identify the at-risk population, followed by health checks tailored to this population. If deemed necessary, behavior change interventions, preventive medical treatment, or a combination of the two may also be included. Various stepwise approaches to health checks have been tested in research studies; however, no long-term effects have been reported [11-15]. In the Danish health care system, health checks are provided to the general population on an opportunistic, nonperiodic, and nontargeted basis. On the basis of a technical feasibility study from 2012 [16], we set up the Tidlig Opsporing og Forebyggelse (TOF; early detection and prevention) project in a partnership with the general practitioners' (GPs) organization and 10 municipalities of the Region of Southern Denmark [17]. Over a period of 2 years, we developed a stepwise model for systematic and targeted prevention of chronic diseases to be used in the Danish primary care sector. The intervention consisted of a joint intervention and a targeted intervention. The joint intervention was applied to the entire study population in the form of a personal digital health profile. The targeted intervention was only applied to patients who were deemed to potentially benefit from either a health check at their GP or lifestyle coaching provided by the municipal health center. Patients at high risk of a chronic disease were identified using validated risk algorithms for COPD, T2DM, and CVD and were offered a health check at their GP in the form of a medical examination and a health dialogue. Patients with health risk behavior included patients who were not at high risk as determined by the risk algorithms but who engaged in one or more health-risk behaviors such as smoking, high-risk alcohol consumption, poor dietary habits, sedentary behavior, and/or a body mass index above 35. This cohort was offered a short 15-min telephone-based health dialogue with a health professional from the municipal health center. For patients with limited health capabilities, the initial telephone-based health dialogue could be followed up by a 1-hour face-to-face health dialogue. If deemed necessary, the targeted intervention would be complemented by further behavior change intervention or preventive medical treatment. Patients already diagnosed with hypertension, hypercholesterolemia, T2DM, CVD, or COPD by the GP, or who displayed no health risk behaviors were only offered the joint intervention. In line with the Medical Research Council's recommendations for complex interventions, we tested the acceptability, feasibility, and short-term effects of the intervention in a pilot study in 2 municipalities between April and December 2016 [18].

A novel feature of the pilot study was the use of a personal digital mailbox for recruiting participants. A personal digital mailbox is a secure digital mailbox provided by the Danish public authorities. It is accessed either via a webpage or an app developed for all major operating systems. The digital mailbox is secured by a national 2-phased log-in system (NemID) and is used by all public authorities as well as an increasing number of private companies such as banks and insurance companies. Beyond being both safe and secure, it is a low-cost, quick, and easy way to reach Danish residents [19]. Permanent residents of Denmark are obliged by law to have a digital mailbox and are expected to check it regularly. Short message service text message and mail reminders are optional. Opting out is only possible in special cases, mainly in the event of low information and technology literacy (usually age-related) or cognitive impairment. A total of 90% of the entire population in Denmark and 95% of the target population have a digital mailbox (May 2016) [20].

Objective

This study reports on the association between the rates of acceptance of 2 digital invitations sent to a personal digital mailbox and sociodemographic determinants, medical treatment, and health care usage in a stepwise primary care model for the prevention of chronic diseases.

Methods

Design

We conducted a cross-sectional analysis of the rates of acceptance of 2 digital invitations sent to residents from 2 Danish municipalities randomly selected to take part in a pilot study of the TOF project (NCT02797392).

Population

The target population consisted of citizens born between 1957 and 1986 and residing in Haderslev or Varde, 2 rural municipalities located in the southern part of Jutland, Denmark. The population of both municipalities totals 106,081 citizens (2015).

Setting

The Danish health care system comprises a strong, publicly funded primary care sector, which includes municipal health centers and GP clinics [21]. GPs operate a patient list system, with an average of 1600 patients per GP. On average, 2 GPs work in a given clinic. Municipal health centers serve the entire population with primary prevention such as smoking cessation and dietary advice, whereas GPs manage and coordinate

secondary prevention, including treatment for hypertension, hypercholesterolemia, and diabetes.

Recruitment Procedure

In January 2016, all 66 GPs residing in the municipalities of Varde and Haderslev received a written invitation, with an enclosed project agreement form and a prepaid return envelope. Nonresponse was followed up by a telephone call to the GP. Using the Regional Primary Care Administrative System (KMD Sygesikring), the regional health authorities identified a source population of 200 randomly selected patients extracted from each of the participating GPs' patient lists. From the source population, we excluded patients having either no digital mailbox or residing outside the municipalities of Varde or Haderslev.

Participants were recruited using 2 digital invitations sent to their personal digital mailbox (Multimedia Appendix 1). Both digital invitations consisted of a 1-page PDF file in Danish and included a highly visible hyperlink to a Web-based digital support system, on which participants would provide both consent and access to the personal digital health profile [22]. The first digital invitation was sent out in April 2016, with the aim of obtaining consent to participate in the study and to access specific information from their GP's electronic patient record (EPR) system, including International Classification of Primary Care, 2nd edition) codes for diagnoses and anatomical therapeutic chemical (ATC) codes for medical prescriptions. The second digital invitation was aimed at providing participants with a digital health profile. This invitation was sent out in September 2016 to participating patients who were registered with the same GP as when they consented and who still resided in the municipalities of Varde or Haderslev. A nonresponse triggered up to 2 reminders 1 week apart.

Outcomes

In this paper, we report on both the consent to take part in the study and the uptake of the personal digital health profile. The main outcome relates to the acceptance of the first digital invitation and is operationalized in terms of consent or nonconsent to take part in the study. Consent is defined as the provision of informed consent to participate in the study; nonconsent includes both nonresponse and active nonconsent. The second outcome relates to the acceptance of the second digital invitation and is operationalized in terms of uptake or nonuptake of the personal digital health profile. Uptake is defined as patients who gave their active consent and received a personal digital health profile. The results from the second digital invitation are presented in Multimedia Appendices 2 and 3.

Variables

Registry variables for the entire study population were retrieved from the administrative registry and Statistics Denmark and linked with the patients' Danish Personal Identification numbers. EPR information was retrieved directly from participating GPs' EPR systems and related purely to consenting patients (Table 1). All participants were pseudonymized when linking project data and national registers from Statistics Denmark.

Age was categorized in 10-year age groups. Country of origin was retrieved for the year 2016 and categorized as Danish, Western, or non-Western origin. Cohabitational status was retrieved for the year 2016 and categorized as cohabiting or single. Highest attainable educational level was retrieved for October 2015 and categorized as secondary school, high school, vocational education, or higher education. Occupation was retrieved for November 2014 and categorized according to the Organisation for Economic Co-operation and Development equivalence scales into 5 groups: employed, self-employed, unemployed or on benefits, social welfare recipients, or other [30]. The distinction between unemployment benefits and social welfare is that unemployment benefits are accessible to citizens who have been unemployed for less than 2 years and who are members of a voluntary unemployment benefit fund. Social welfare benefits are for all other unemployed persons who can take up a job. Others represent, for example, nonworking persons from a family that relies on 1 income only. Family income was retrieved for 2013, 2014, and 2015, defined by the mean annual net income of the household, and categorized into quartiles. "Partner in project" describes whether your partner (if cohabiting) participated as well. Partner in project is categorized in a binary yes or no variable.

Information on prescriptions and diagnoses was combined as a proxy for medical treatment (Table 2). Prescriptions were retrieved for the period from May 2014 to April 2016 as ATC codes. We chose a 2-year period, as prescriptions may be filed up to 2 years after their date of issue. International Classification of Diseases 10th edition (ICD-10) codes were retrieved for the period from January 2013 to April 2016. Medical treatment was defined as either registered with an ATC code, ICD-10 code, or both during the periods specified above.

Table 1. Analyses of associations between patient characteristics and acceptance.

Variable	First digital invitation	Second digital invitation	
Presentation of results	Results section	Multimedia Appendix 2	Multimedia Appendix 3
Denominator	Study population	Study population	Consent to the first digital invitation
Outcome variable	Consent or nonconsent to the first digital invitation	Uptake or nonuptake of the second digital invitation	Uptake or nonuptake of the second digital invitation
Exposures			
Sociodemographics	Age[a,b]	Age[a,b]	Age[a,b]
	Sex[a,b]	Sex[a,b]	Sex[a,b]
	Country of origin[b]	Country of origin	Country of origin
	Highest educational attainment[c]	Highest educational attainment[c]	Highest educational attainment[c]
	Occupational status[d]	Occupational status[d]	Occupational status[d]
	Family income[e]	Family income[e]	Family income[e]
	Cohabitational status[b]	Cohabitational status[b]	Cohabitational status[b]
	Partner consent[f]	Partner consent[f]	Partner consent[f]
Medical treatment	Prescriptions from primary care or hospitals (anatomical therapeutic chemical [ATC] codes)[g]	Prescriptions from primary care or hospitals (ATC codes)[g]	Prescriptions from primary care or hospitals (ATC codes)[g]
	Hospital discharge diagnoses (International Classification of Diseases 10th edition [ICD-10] codes)[h]	Hospital discharge diagnoses (ICD-10 codes)[h]	Hospital discharge diagnoses (ICD-10 codes)[h]
Health care usage	Administrative primary care codes[i,j]	Administrative primary care codes[i,j]	Administrative primary care codes[i,j]; health checks[i,j]

[a]Danish National Administrative Primary Care System (Praksys).
[b]Danish Civil Registration System [23].
[c]Danish Education Register [24].
[d]Danish Registers of Labour Market Affiliation [25].
[e]Danish Registers of Personal Income and Transfer Payments [26].
[f]Questionnaire data.
[g]Danish National Prescription Registry [27].
[h]Danish National Patient Registry [28].
[i]General practitioners' electronic patient record.
[j]Danish National Health Service Register [29].

Table 2. International Classification of Diseases 10th edition codes and anatomical therapeutic chemical classification codes used to define when a medical condition of chronic obstructive pulmonary disease, cardiovascular disease, or diabetes mellitus had been registered.

Medical condition	ICD-10[a] codes registered from January 2013 to April 2016	ATC[b] therapeutic codes for prescribed medicine registered from May 2014 to April 2016
COPD[c]	J44	R03AC18, R03AC19, R03AL03, R03AL04, R03AL05, R03BB04, R03BB05, and R03BB06
CVD[d]	I1-I7 and E78 (except: I0, I16, I60, I73, and I78)	C (except: C01CA and C05)
Diabetes mellitus	E10-E13	A10

[a]ICD-10: International Classification of Diseases 10th edition.
[b]ATC: anatomical therapeutic chemical.
[c]COPD: chronic obstructive pulmonary disease.
[d]CVD: cardiovascular disease.

Health care usage was determined from administrative codes registered by the GP and retrieved for the period from May 2013 to April 2016. To this end, we also examined EPR information on laboratory test results. Administrative codes were used to extract frequent GP attenders, GP attenders, and usage of specific administrative codes pertaining to laboratory tests and preventive consultations. Frequent attenders referred to the top 10% of patients, who on average contacted the GP the most—either in person or by phone—during the 3 years from May 2013 to April 2016 [31]. GP attenders were defined as patients having contacted their GP during the 2-year period from May 2014 to April 2016. Laboratory tests comprised blood samples (administrative codes 2101 and 2601), peak flow tests (7183), spirometries (7113), and home blood pressure monitoring (2146) during the period from May 2014 to April 2016. Furthermore, we retrieved information on specific preventive consultations (0120)—that is, special consultations for coaching patients with health risk behaviors and patients diagnosed with a chronic disease. To determine if a patient had received a preventive health check within a period of 2 years before consenting to the study (May 2014 to May 2016), we retrieved blood pressure (systolic and diastolic blood pressure), lung function (forced expiratory volume [FEV_1], forced vital capacity [FVC], and FEV_1/FVC), glycated hemoglobin, and lipids (total cholesterol, high-density lipoprotein, and low-density lipoprotein) measurements from the GPs' EPR systems. A health check was defined as having had 2 or more of the above-mentioned values measured in the same consultation.

Analysis

Statistical associations are presented as crude figures, age- and sex-adjusted figures, and as models minimally adjusted for known confounders. Poisson regression with robust variance error was used rather than logistic regression to obtain incidence rate ratios (IRR). The binary outcome variable of consent or nonconsent was interpreted as a continuous variable with the only counts being 0 or 1. The minimally adjusted model was developed from a causal direct acyclic graph (DAG), built on the current evidence of the determinants of attendance at health checks [32]. Attendance at health checks was the outcome variable of the DAG. Significance level was set at $P<.05$.

A data-driven chi-square automatic interaction detection (CHAID) method was used to generate a decision tree analysis to identify interactions and a hierarchy of variables predicting the chosen outcome variable (root node) [33]. Using chi-square tests of interdependence, a CHAID analysis clusters categories within each predictor variable to determine what predictor variables are associated with the outcome. Subsequently, the predictor variables associated with the outcome are ordered hierarchically. The specific variable order is determined by the Bonferroni *P* value of each variable such that the predictor variable with the smallest *P* value (strongest association) is placed at the top of the hierarchy (parent node). The minimum number of observations for each split in the decision tree was set at 500 (child node) and 200 for each node (terminal node). In the analysis of the acceptance of the second digital invitation, each split in the decision tree was set to 100 (child node) and 20 for each node (terminal node) because of the limited number of observations.

Statistical analysis was performed on secure servers at Statistics Denmark using Stata 14 (Statacorp).

Ethics Approval and Consent to Participate

The study was approved by the Danish Data Protection Agency (J.Number 2015-57-0008) and registered at Clinical Trial Gov (Unique Protocol ID: TOFpilot2016). According to Danish regulations (Act on Research Ethics Review of Health Research Projects [section 14,2]), this study does not need approval from a health research ethics committee as no research on human tissue or other biological material is performed. The study complies with the Helsinki Declaration, with informed consent to study participation and to disclosure of data from the GPs' EPR obtained from all participants.

Results

Recruitment and Overall Uptake

Of the 68 GPs residing in the 2 municipalities, 47 GPs from 18 clinics agreed to participate in the study (Figure 1). This provided us with a source population of 9400 patients. However, a total of 586 patients did not meet the inclusion criteria, which is why only 8814 received the first invitation. Initially, a total of 3587 patients consented to participate, but among them, 30 patients moved from the municipality to a nonparticipating GP, and 12 withdrew their consent after receiving the second invitation. This resulted in 3545 active consenters from the first round of invitations (Multimedia Appendix 2). Of the patients who accepted the first digital invitation (n=3545), 75.06% (2661/3545) also accepted the second digital invitation (Multimedia Appendix 3; Table 3).

Figure 1. Flow diagram from source population to study population. GP: general practitioner.

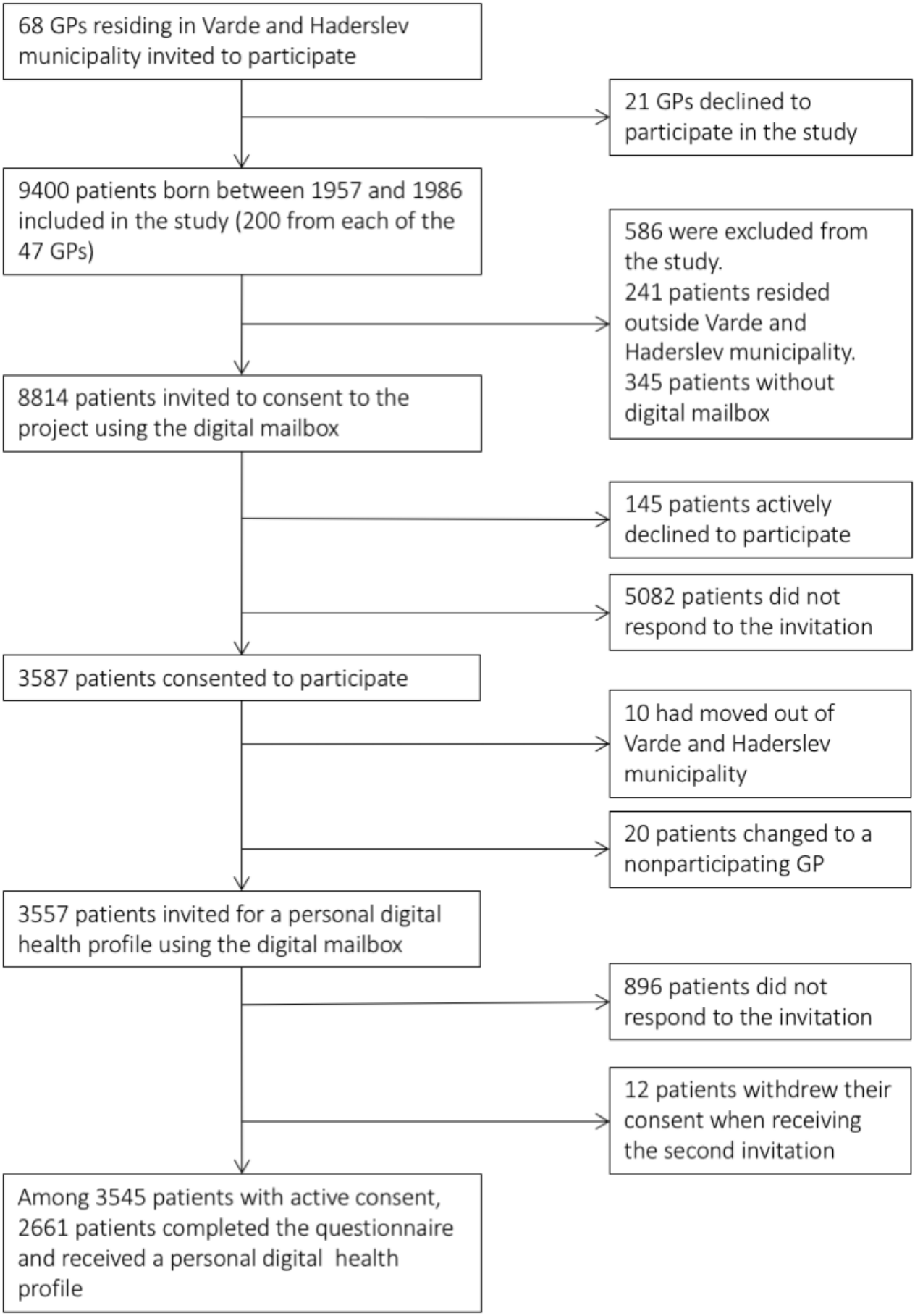


Table 3. Descriptive analysis of determinants of the acceptance of the first digital invitation.

Determinants	Consenters (N=3545), n (%)	Nonconsenters (N=5269), n (%)	Total (N=8814), n (%)	Missing, n (%)
Demography[a]				
Age (years)				
29-39	732 (20.64)	1921 (36.45)	2653 (30.09)	0 (0.00)
40-49	1151 (32.46)	1875 (35.58)	3026 (34.33)	0 (0.00)
50-60	1662 (46.88)	1473 (27.95)	3135 (35.56)	0 (0.00)
Sex				
Male	1590 (44.85)	2845 (53.99)	4379 (49.68)	0 (0.00)
Female	1955 (55.14)	2424 (46.00)	4435 (50.31)	0 (0.00)
Country of origin				
Denmark	3385 (95.48)	4446 (84.38)	7831 (88.44)	18 (0.20)
Western	91 (2.56)	458 (8.69)	549 (6.22)	18 (0.20)
Non-Western	69 (1.94)	347 (6.58)	416 (4.71)	18 (0.20)
Cohabitation				
Single	726 (20.47)	1516 (28.77)	2242 (25.43)	18 (0.20)
Cohabiting	2819 (79.52)	3735 (70.88)	6554 (74.35)	18 (0.20)
Partner in project				
Yes	2037 (57.46)	2715 (51.52)	4752 (53.91)	18 (0.20)
No	1508 (42.53)	2536 (48.13)	4044 (45.88)	18 (0.20)
Socioeconomy				
Educational attainment				
Secondary school	513 (14.47)	1194 (22.66)	1707 (19.36)	583 (6.61)
High school	143 (4.103)	213 (4.04)	356 (4.03)	583 (6.61)
Vocational education	1604 (45.24)	2199 (41.73)	3803 (46.14)	583 (6.61)
Higher education	1216 (34.30)	1149 (21.80)	2365 (26.83)	583 (6.61)
Employment status				
Employed	2891 (81.55)	3719 (70.58)	6610 (74.99)	105 (1.19)
Self-employed	170 (4.79)	260 (4.93)	430 (4.87)	105 (1.19)
Benefits	88 (2.48)	184 (3.49)	272 (3.08)	105 (1.19)
Social welfare	340 (9.59)	806 (15.29)	1146 (13.00)	105 (1.19)
Other	51 (1.43)	200 (3.79)	251 (2.84)	105 (1.19)
Family income				
Low	559 (15.76)	1488 (28.24)	2047 (23.22)	130 (1.47)
Middle-low	806 (22.73)	1341 (25.45)	2147 (24.35)	130 (1.47)
Middle-high	989 (27.89)	1246 (24.64)	2235 (25.35)	130 (1.47)
High	1183 (33.4)	1072 (20.34)	2255 (25.58)	130 (1.47)
Medical treatment[b]				
Prescriptions and diagnoses				
Treatment	763 (21.52)	973 (18.46)	1736 (19.69)	0 (0.00)
No treatment	2782 (78.47)	4296 (81.53)	7078 (80.30)	0 (0.00)
Health care usage[c]				
Attendance to general practitioner (GP)				

Determinants	Consenters (N=3545), n (%)	Nonconsenters (N=5269), n (%)	Total (N=8814), n (%)	Missing, n (%)
Yes	3173 (89.50)	4372 (82.97)	7545 (85.60)	0 (0.00)
No	372 (10.49)	897 (17.02)	1269 (14.39)	0 (0.00)
Frequent attender				
Yes	368 (10.38)	584 (11.08)	952 (10.80)	0 (0.00)
No	3177 (89.61)	4685 (88.91)	7862 (89.19)	0 (0.00)
Laboratory tests at GP				
Yes	2053 (57.91)	2471 (46.89)	4524 (51.32)	0 (0.00)
No	1492 (42.08)	2798 (53.10)	4290 (48.67)	0 (0.00)
Preventive consultation at GP				
Yes	431 (12.15)	516 (9.79)	947 (10.74)	0 (0.00)
No	3114 (87.84)	4753 (90.20)	7867 (89.25)	0 (0.00)

[a]Social registries.

[b]Anatomical therapeutic chemical codes and International Classification of Diseases 10th edition codes related to diabetes, cardiovascular disease, and chronic obstructive pulmonary disease.

[c]Administrative codes from the general practitioner.

Acceptance of the First Digital Invitation

The Poisson regressions showed that a higher rate of acceptance of the first digital invitation was associated with sociodemographic factors, including higher age, income, and educational attainment (Table 4). A higher rate of acceptance was also associated with being female, employed, born in Denmark, and cohabiting. Patients not diagnosed with or in treatment for T2DM, CVD, or COPD were more likely to accept the first digital invitation than patients in treatment. Similarly, the acceptance rate was higher among patients who had seen their GP or who had registered 1 or more laboratory tests at their GP within 2 years of giving consent. We found no association between the likelihood of accepting the first digital invitation and the frequency of GP appointments, being registered with a preventive consultation, or having a partner that also consented to the study.

The CHAID analysis showed that age was the strongest predictor of accepting the first digital invitation followed by the educational attainment in patients below the age of 50 years and income in patients above the age of 50 years (Figure 2). The CHAID showed large subgroup differences in acceptance rates. Of patients below the age of 40 years, with secondary school as the highest educational attainment, 15.92% accepted the first digital invitation. By contrast, the acceptance rate among patients above the age of 50 years, with high income and with at least a bachelor-level education was 68.58 %.

Table 4. Analysis of associations between acceptance of the first digital invitation and sociodemographic determinants, medical treatment, and health care usage.

Determinants	Sample size (N)	Model 1 (crude)		Model 2 (adjusted for age and sex)		Model 3 (minimally adjusted)	
		IRR[a] (95% CI)	*P* value	IRR (95% CI)	*P* value	IRR (95% CI)	*P* value
Age[b] (years)							
29-39	2653	1 (0)	__[c]	1 (0)	—	1 (0)	—
40-49	3026	1.08 (1.06-1.10)	.001	1.08 (1.06-1.10)	.001	1.08 (1.06-1.10)	.001
50-60	3135	1.20 (1.18-1.22)	.001	1.20 (1.18-1.22)	.001	1.20 (1.18-1.22)	.001
Sex[b]							
Female	4435	1 (0)	—	1 (0)	—	1 (0)	—
Male	4379	0.94 (0.93-0.95)	.001	0.94 (0.93-0.95)	.001	0.94 (0.93-0.95)	.001
Country of origin[b]							
Denmark	7831	1 (0)	—	1 (0)	—	1 (0)	—
Western	549	0.81 (0.79-0.84)	.001	0.84 (0.81-0.86)	.001	0.81 (0.79-0.84)	.001
Non-Western	416	0.81 (0.79-0.84)	.001	0.84 (0.81-0.86)	.001	0.81 (0.79-0.84)	.001
Cohabitation[d]							
Single	2242	1 (0)	—	1 (0)	—	1 (0)	—
Cohabiting	6554	1.08 (1.06-01.10)	.001	1.07 (1.05-01.08)	.001	1.05 (1.03-01.07)	.001
Partner in project[e]							
Yes	4752	1 (0)	—	1 (0)	—	1 (0)	—
No	4044	0.96 (0.95-0.98)	.001	0.96 (0.95-0.98)	.001	1.00 (0.99-1.02)	.70
Educational attainment[f]							
Secondary school	1707	1 (0)	—	1 (0)	—	1 (0)	—
High school	356	1.08 (1.04-1.12)	.001	1.09 (1.05-1.13)	.001	1.09 (1.05-1.14)	.001
Vocational education	3803	1.09 (1.07-1.12)	.001	1.10 (1.08-1.12)	.001	1.10 (1.08-1.12)	.001
Higher education	2365	1.16 (1.14-1.19)	.001	1.17 (1.14-1.19)	.001	1.16 (1.14-1.19)	.001
Employment status[d]							
Employed	6610	1 (0)	—	1 (0)	—	1 (0)	—
Self-employed	430	0.97 (0.94-1.00)	.089	0.96 (0.93-1.00)	.03	0.97 (0.94-1.00)	.05
Benefits	272	0.92 (0.88-0.96)	.001	0.93 (0.89-0.97)	.001	0.95 (0.91-1.00)	.03
Social welfare	1146	0.90 (0.88-0.92)	.001	0.90 (0.88-0.92)	.001	0.94 (0.92-0.96)	.001
Other	251	0.84 (0.80-0.87)	.001	0.85 (0.82-0.89)	.001	0.92 (0.87-0.97)	.005
Family income[d]							
Low	2047	1 (0)	—	1 (0)	—	1 (0)	—
Middle-low	2147	1.08 (1.06-1.10)	.001	1.07 (1.05-1.10)	.001	1.05 (1.02-1.07)	.001
Middle-high	2235	1.13 (1.11-1.16)	.001	1.11 (1.09-1.14)	.001	1.06 (1.04-1.09)	.001
High	2255	1.20 (1.17-1.22)	.001	1.14 (1.12-1.17)	.001	1.08 (1.05-1.10)	.001
Prescriptions and diagnoses[d]							
Treatment	1736	1 (0)	—	1 (0)	—	1 (0)	—
No treatment	7078	0.97 (0.95-0.99)	.001	1.02 (1.00-1.04)	.02	1.02 (1.00-1.04)	.04
Attendance at general practitioner (GP)[d]							
Yes	7545	1 (0)	—	1 (0)	—	1 (0)	—

Determinants	Sample size (N)	Model 1 (crude)		Model 2 (adjusted for age and sex)		Model 3 (minimally adjusted)	
		IRR[a] (95% CI)	*P* value	IRR (95% CI)	*P* value	IRR (95% CI)	*P* value
No	1269	0.91 (0.89-0.93)	.001	0.93 (0.91-0.95)	.001	0.95 (0.93-0.98)	.001
Frequent attender to GP[d]							
Yes	952	1 (0)	—	1 (0)	—	1 (0)	—
No	7862	1.01 (0.99-1.04)	.30	1.03 (1.01-1.06)	.007	1.02 (1.00-1.05)	.07
Laboratory tests at GP[d]							
Yes	4524	1 (0)	—	1 (0)	—	1 (0)	—
No	4290	0.93 (0.91-0.94)	.001	0.95 (0.94-0.97)	.001	0.95 (0.94-0.97)	.001
Preventive consultation at GP[d]							
Yes	947	1 (0)	—	1 (0)	—	1 (0)	—
No	7867	0.96 (0.94-0.98)	.001	1.00 (0.98-1.02)	.94	0.99 (0.97-1.02)	.66

[a]Incidence rate ratio.

[b]Model 3 adjustments: no adjustments.

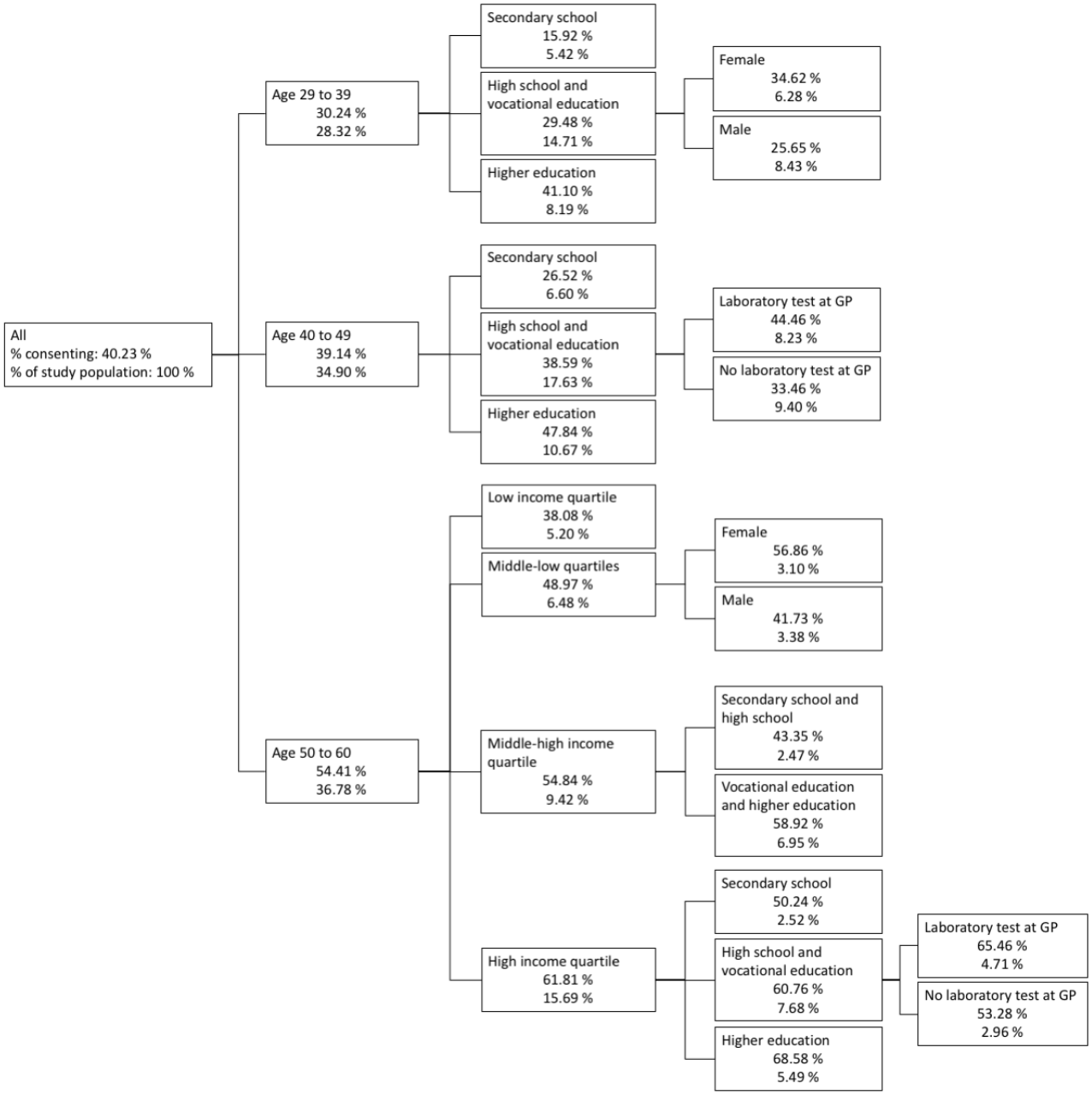
[c]Reference category.

[d]Model 3 adjustments: age, sex, country of origin, and education.

[e]Model 3 adjustments: cohabitation.

[f]Model 3 adjustments: age, sex, and country of origin.

Figure 2. Chi-square automatic interaction detection analysis of the acceptance of the first digital invitation. GP: general practitioner.



Acceptance of the Second Digital Invitation

In the entire study population, the rate of acceptance of the second digital invitation showed associations similar to the ones in the analysis of the first digital invitation. The only differences were higher rates of acceptance among frequent GP attenders and no association with attendance or nonattendance to the GP. The similarity also applied to the CHAID analysis, in which age was shown to be the strongest predictor, followed by educational attainment in the age below 50-years bracket and income in the age above 50-years bracket. The CHAID analysis showed a rate of acceptance of 8.37% in patients below the age of 40, with secondary school as their highest educational attainment and income below 50% of the median. By contrast, among patients aged more than 50 years, with an income of more than 50% above the median and at least a bachelor-level education, 60.40% accepted the second digital invitation.

Acceptance of the Second Digital Invitation Among Consenters (N=3545)

Among the patients who accepted the first digital invitation, the Poisson regressions indicated associations between the rate of acceptance of the second digital invitation and most sociodemographic variables. The rate of acceptance increased with age, educational attainment, and income. Being female, employed, and born in Denmark also correlated positively with rates of acceptance. We saw higher rates of acceptance among patients who had not seen their GP within 2 years before consenting to the study. No other variables describing health status or health care usage showed an association with rates of acceptance—that includes having received a health check during a period of 2 years before consenting to the study.

The CHAID analysis showed that age was the strongest predictor of acceptance of the second digital invitation among those who

accepted the first digital invitation and the only predictor in patients younger than 50 years. In patients older than 50 years, income followed age as the second strongest predictor. Specifically, when it came to the patients older than 50 years, with a middle-low or middle-high income, the CHAID analysis revealed that nonattendance at the GP and not receiving medical treatment during a period of 2 years before consenting were both strong predictors of whether or not participants accepted the second digital invitation.

Discussion

Main Findings

In this study, 40.22% of our sample accepted the first digital invitation and 30.19% accepted both digital invitations. That is, among those who accepted the first digital invitation, 75.06% also accepted the second. The rates of acceptance of both digital invitations were higher among women, elderly patients, patients of higher SES, and patients not diagnosed with or in treatment for T2DM, COPD, or CVD. Patients who had seen their GP within the past 2 years were also more likely to accept the digital invitations. The frequency of GP appointments, registering for a preventive consultation, or having a partner who had accepted the first digital invitation showed no association with the acceptance of either invitation. In the subpopulation of patients (N=2661) who accepted the first digital invitation, women, patients of relatively high age and SES, and patients who had not seen their GP for a period of 2 years before giving their consent showed a higher rate of acceptance of the second invitation. No other health care usage, including having had a health check within the previous 2 years or being diagnosed with or in treatment for T2DM, COPD, or CVD showed any association with the rate of acceptance.

Low patient uptake of stepwise models for preventing chronic diseases seems to be the current norm. This trend may be attributed to a combination of a recent overall increase in the use of preventive health checks in primary care and a decrease in response rates to research studies in general [34,35]. A recent Dutch study of a stepwise prevention model showed an uptake rate of 29% in patients aged 45 to 70 years, whereas an Australian study had an initial uptake rate of 31% in patients aged 40 to 64 years [11,36]. A feasibility study to this study showed an uptake rate of 63% using paper-based invitations, with a link to a Web-based questionnaire and an enclosed hard copy questionnaire and return envelope [37]. Two other Danish studies of stepwise models showed uptake rates of 55% in a general population aged 30 to 49 years and 30% in a population of social housing residents aged 45 to 70 years [38,39]. Both Danish studies used a proactive approach by which paper-based invitations indicated a prebooked time and date—a method which has been shown to garner increased response rates in a previous study [40]. Furthermore, the associations between acceptance of digital invitations and socioeconomic determinants are in line with the evidence from other Danish and European studies of health checks [6,38,41]. However, the CHAID analyses showed that the differences in rates of acceptance among the SES groups in our study are larger than those observed in the 2 Danish studies mentioned above [38,39]. The evidence of the association between uptake of stepwise models and medical treatment or health care usage is scarce and largely inconclusive. This is presumably because of a general lack of health and health care information on nonresponders. Nonetheless, the results of medical treatment and health care usage in this study differ somewhat from previous studies. We show slightly higher rates of acceptance of the digital invitations among patients not diagnosed with or in treatment for T2DM, COPD, or CVD. This is in line with a comparable Danish study [38], whereas other studies report either higher uptake among patients with chronic diseases [42,43] or no association [44,45]. The inconsistency in results could be explained by different definitions of medical treatment. Similarly, different definitions of preventive services could explain why we see no association between acceptance of the digital invitations and use of preventive consultations at the GP, whereas other studies suggest an association [38]. We saw no association between acceptance of either invitation and having had a health check during a period of 2 years leading up to consent; however, other studies have consistently found that prior use of health checks seems to increase the likelihood of getting another health check [46-48]. Interestingly, we found an association between not having had a GP appointment during the previous 2 years and a higher rate of acceptance of the second digital invitation among those who also accepted the first digital invitation. This may suggest that true compliers to the study (ie, patients who would not have taken up the offer had they not been invited) are more likely to also accept the second digital invitation than always-users (ie, patients who always respond to invitations to participate in preventive services) [49]. As we saw no association between the rates of acceptance of the second digital invitation and other variables of health care utilization, this result should be interpreted with great care and examined further in future studies.

Efficacy of Digital Recruitment

To our knowledge, this is the first study to report on the rates of acceptance of digital invitations to participate in a stepwise model for prevention of chronic diseases. Moreover, it is most likely the first to report on digital recruitment to a health intervention using digital invitations. We cannot establish whether rates of acceptance would have been different if recruitment had been paper-based, as we did not include a random subpopulation, which received paper-based invitations. To our knowledge, comparisons of digital and paper-based invitations sent by regular mail have only been reported once in a randomized study by Ebert et al [19]. This study showed that 50- to 59-year-old responders to the digital invitation were more likely to be of higher SES than their counterparts who responded to paper invitations. However, no differences were seen in those aged 30 to 39 years. The overall rate of acceptance of digital invitations and paper-based invitations was comparable. The rate of acceptance of digital invitations in this study and the results reported by Ebert et al may suggest slightly lower overall acceptance rates and slightly larger SES differences when using digital invitations. However, the rate of acceptance of digital invitations combined with Web-based data collection may resemble the emergence of the combined paper-based invitation and Web-based data-collection approach,

where the advent of Web-based data collection methods precipitated an initial drop in the rate of acceptance [34].

In addition, digital invitations sent to personal digital mailboxes seem to be an especially suitable and low-cost recruitment method for patients of high SES. It is well known that risk factors of chronic lifestyle-related diseases are clustered in low SES populations [50]. To generate population health effects, stepwise models for the prevention of chronic diseases may have to employ other low-tech recruitment approaches, which complement digital invitations. Results from the health check program of the British National Health Service indicate that uptake may increase over time, with a clear focus on the hardest-to-reach populations [51]. Thus, it would appear that the lower uptake among patients of lower SES found in this as well as in many other studies can be eliminated, or even inverted, by a focused recruitment effort aimed at deprived communities and outreach services [52,53]. However, at present, digital recruitment is only applicable in a few countries, among others the Nordic countries of Denmark, Norway, and Sweden. When digital mail gets more widespread, the results from the described recruitment procedure and intervention may be well applicable in other settings as well.

Nonetheless, more studies of digital invitations are needed to determine if the acceptance rates seen in this study and in the study by Ebert et al could be expected from future studies as well. Similarly, more research is needed to determine whether a multimodal recruitment approach, including digital invitations to personal digital mailboxes will reach other subpopulations more effectively than digital invitations only.

Strengths and Limitations

The main strengths of this study relate to the high validity of the registries of Statistics Denmark and the random sampling of patients from a large number of GP clinics. Especially, the health and social registries are of high quality, with few missing cases and up-to-date information that has been registered either immediately before study commencement or during another specified period before study commencement [54]. Thus, the impact is most likely negligent because of the strong association between acceptance of the digital invitations and SES. All direct contacts with the GP are most likely both valid and complete as this type of information is automatically registered onto patients' personal health insurance cards. Other administrative data from the primary care sector on specific tasks performed as part of a consultation, such as taking a blood sample, a spirometry, or having a preventive consultation, may be incomplete and more prone to human error and should as such be interpreted with care. Another strength is the combination of DAG and CHAID analyses to establish a both theory-driven and data-driven analytical approach. The DAG established adjustments to the Poisson regression models based on a theoretical and evidence-based causal model. The CHAID used the data from this study and identified the strongest predictors of attendance. However, residual confounding and collider bias cannot be eliminated in the regressions because of both a rather complex causal model, conditional independence between exposures, and the unavailability of a number of exposures—especially health-risk behaviors as well as cognitive and psychological parameters [6]. Finally, the invitations sent to the digital mailbox and the digital support system were in Danish language only, which may have had a negative impact on the rate of acceptance among people originating from outside of Denmark.

Conclusions

To our knowledge, this is the first study to report on the rates of acceptance of digital invitations to participate in a stepwise model for prevention of chronic diseases. We show acceptance rates of 40% for a first digital invitation and 30% for a second digital invitation, the rates being higher among women, elderly patients, and patients of higher SES; patients not diagnosed with, or in treatment for T2DM, COPD, and/or CVD; and patients having attended the GP within a period of 2 years before consent. A total of 75% of those who accepted the first digital invitation also accepted the second. On one hand, the 2 digital invitations seem to deepen the sociodemographic differences in acceptance compared with a single digital invitation; on the other hand, patients who had not consulted a GP during a period of 2 years before the study, and who were not receiving medical treatment, showed a significantly higher rate of acceptance of the second invitation when it came to those who had accepted the first digital invitation. This suggests that compliers are more prone to accepting the second digital invitation than always-users are. In future studies, multimodal recruitment approaches, which complement digital invitations, are warranted to increase the rates of acceptance among harder-to-reach subpopulations.

Acknowledgments

The project received funding from the Region of Southern Denmark and the Danish Ministry of Health. The funders had no role in the analysis or in the interpretation of the data, writing of the manuscript, or the decision to submit it for publication.

Authors' Contributions

All authors participated in the design of the study. LBL analyzed and interpreted the patient data and was the main contributor to the manuscript. All authors contributed to the interpretation of patient data and the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Invitations (in Danish).

[PDF File (Adobe PDF File), 310KB - jmir_v21i1e11658_app1.pdf]

Multimedia Appendix 2

Data on the acceptance of the second digital invitation in the entire study population.

[PDF File (Adobe PDF File), 290KB - jmir_v21i1e11658_app2.pdf]

Multimedia Appendix 3

Data on the acceptance of the second digital invitation among those who accepted the first digital invitation.

[PDF File (Adobe PDF File), 187KB - jmir_v21i1e11658_app3.pdf]

References

1. Krogsbøll LT, Jørgensen KJ, Grønhøj Larsen C, Gøtzsche PC. General health checks in adults for reducing morbidity and mortality from disease: Cochrane systematic review and meta-analysis. Br Med J 2012 Nov 20;345:e7191 [FREE Full text] [doi: 10.1136/bmj.e7191] [Medline: 23169868]
2. Si S, Moss J, Sullivan T, Newton S, Stocks N. Effectiveness of general practice-based health checks: a systematic review and meta-analysis. Br J Gen Pract 2014 Jan;64(618):e47-e53 [FREE Full text] [doi: 10.3399/bjgp14X676456] [Medline: 24567582]
3. Prochazka A, Caverly T. General health checks in adults for reducing morbidity and mortality from disease: summary review of primary findings and conclusions. JAMA Intern Med 2013 Mar 11;173(5):371-372. [doi: 10.1001/jamainternmed.2013.3187] [Medline: 23318544]
4. Capewell S, Graham H. Will cardiovascular disease prevention widen health inequalities? PLoS Med 2010 Aug 24;7(8):e1000320 [FREE Full text] [doi: 10.1371/journal.pmed.1000320] [Medline: 20811492]
5. Gøtzsche P, Jørgensen K, Krogsbøll L. General health checks don't work. Br Med J 2014;348:g3680. [doi: 10.1136/bmj.g3680]
6. Dryden R, Williams B, McCowan C, Themessl-Huber M. What do we know about who does and does not attend general health checks? Findings from a narrative scoping review. BMC Public Health 2012 Aug 31;12:723 [FREE Full text] [doi: 10.1186/1471-2458-12-723] [Medline: 22938046]
7. Bender A, Jørgensen T, Pisinger C. Is self-selection the main driver of positive interpretations of general health checks? The Inter99 randomized trial. Prev Med 2015 Dec;81:42-48. [doi: 10.1016/j.ypmed.2015.07.004] [Medline: 26190370]
8. Shaw R, Holland C, Pattison H, Cooke R. Patients' perceptions and experiences of cardiovascular disease and diabetes prevention programmes: a systematic review and framework synthesis using the Theoretical Domains Framework. Soc Sci Med 2016 May;156:192-203. [doi: 10.1016/j.socscimed.2016.03.015] [Medline: 27043372]
9. Piepoli M, Hoes A, Agewall S, Albus C, Brotons C, Catapano AM, ESC Scientific Document Group. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: the Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J 2016 Dec 01;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]
10. Choices NHS.: NHS; 2012. NHS Health Check Internet URL: https://www.nhs.uk/conditions/nhs-health-check/ [accessed 2018-11-20] [WebCite Cache ID 744xd88A1]
11. van den Brekel-Dijkstra K, Rengers A, Niessen M, de Wit NJ, Kraaijenhagen R. Personalized prevention approach with use of a web-based cardiovascular risk assessment with tailored lifestyle follow-up in primary care practice--a pilot study. Eur J Prev Cardiol 2016 Mar;23(5):544-551. [doi: 10.1177/2047487315591441] [Medline: 26080811]
12. Marshall T, Westerby P, Chen J, Fairfield M, Harding J, Westerby R, et al. The Sandwell Project: a controlled evaluation of a programme of targeted screening for prevention of cardiovascular disease in primary care. BMC Public Health 2008 Feb 25;8:73 [FREE Full text] [doi: 10.1186/1471-2458-8-73] [Medline: 18298863]
13. Marshall T, Caley M, Hemming K, Gill P, Gale N, Jolly K. Mixed methods evaluation of targeted case finding for cardiovascular disease prevention using a stepped wedged cluster RCT. BMC Public Health 2012;12:908 [FREE Full text] [doi: 10.1186/1471-2458-12-908] [Medline: 23101763]
14. Badenbroek IF, Stol DM, Nielen MM, Hollander M, Kraaijenhagen RA, de Wit GA, et al. Design of the INTEGRATE study: effectiveness and cost-effectiveness of a cardiometabolic risk assessment and treatment program integrated in primary care. BMC Fam Pract 2014 May 09;15:90 [FREE Full text] [doi: 10.1186/1471-2296-15-90] [Medline: 24884779]
15. Maindal H, Støvring H, Sandbaek A. Effectiveness of the population-based Check your health preventive programme conducted in primary care with 4 years follow-up [the CORE trial]: study protocol for a randomised controlled trial. Trials 2014 Aug 29;15:341 [FREE Full text] [doi: 10.1186/1745-6215-15-341] [Medline: 25169211]

16. Bruun Larsen L, Soendergaard J, Halling A, Thilsing T, Thomsen J. A novel approach to population-based risk stratification, comprising individualized lifestyle intervention in Danish general practice to prevent chronic diseases: Results from a feasibility study. Health Informatics J 2017 Dec;23(4):249-259. [doi: 10.1177/1460458216645149] [Medline: 27245672]
17. Larsen LB, Sonderlund AL, Sondergaard J, Thomsen JL, Halling A, Hvidt NC, et al. Targeted prevention in primary care aimed at lifestyle-related diseases: a study protocol for a non-randomised pilot study. BMC Fam Pract 2018;19:124. [doi: 10.1186/s12875-018-0820-8]
18. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. Br Med J 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
19. Ebert JF, Huibers L, Christensen B, Christensen MB. Paper- or web-based questionnaire invitations as a method for data collection: cross-sectional comparative study of differences in response rate, completeness of data, and financial cost. J Med Internet Res 2018 Jan 23;20(1):e24 [FREE Full text] [doi: 10.2196/jmir.8353] [Medline: 29362206]
20. Digitization Agency. 2018. Monthly statistics on access to digital mail in Denmark Internet URL: https://digst.dk/it-loesninger/digital-post/om-loesningen/tal-og-statistik/ [accessed 2018-09-27] [WebCite Cache ID 72jZRhJGB]
21. Pedersen K, Andersen J, Søndergaard J. General practice and primary health care in Denmark. J Am Board Fam Med 2012 Mar;25(Suppl 1):S34-S38 [FREE Full text] [doi: 10.3122/jabfm.2012.02.110216] [Medline: 22403249]
22. tof.sundhedsmappe.dk. Odense: Dansk Almen Medicinsk KvalitetsEnhed (DAK-E) URL: https://tof.sundhedsmappe.dk/ [accessed 2018-09-27]
23. Pedersen C. The Danish civil registration system. Scand J Public Health 2011 Jul;39(7 Suppl):22-25. [doi: 10.1177/1403494810387965] [Medline: 21775345]
24. Jensen V, Rasmussen A. Danish education registers. Scand J Public Health 2011 Jul;39(7 Suppl):91-94. [doi: 10.1177/1403494810394715] [Medline: 21775362]
25. Petersson F, Baadsgaard M, Thygesen L. Danish registers on personal labour market affiliation. Scand J Public Health 2011 Jul;39(7 Suppl):95-98. [doi: 10.1177/1403494811408483] [Medline: 21775363]
26. Baadsgaard M, Quitzau J. Danish registers on personal income and transfer payments. Scand J Public Health 2011 Jul;39(7 Suppl):103-105. [doi: 10.1177/1403494811405098] [Medline: 21775365]
27. Pottegård A, Schmidt S, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data resource profile: the Danish national prescription registry. Int J Epidemiol 2017 Dec 01;46(3):798-798f [FREE Full text] [doi: 10.1093/ije/dyw213] [Medline: 27789670]
28. Schmidt M, Schmidt S, Sandegaard J, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol 2015;7:449-490 [FREE Full text] [doi: 10.2147/CLEP.S91125] [Medline: 26604824]
29. Andersen J, Olivarius N, Krasnik A. The Danish national health service register. Scand J Public Health 2011 Jul;39(7 Suppl):34-37. [doi: 10.1177/1403494810394718] [Medline: 21775348]
30. OECD. What are the equivalence scales Internet URL: http://www.oecd.org/els/soc/OECD-Note-EquivalenceScales.pdf [accessed 2018-11-20] [WebCite Cache ID 7450EgjQd]
31. Vedsted P, Fink P, Sørensen HT, Olesen F. Physical, mental and social factors associated with frequent attendance in Danish general practice. A population-based cross-sectional study. Soc Sci Med 2004 Aug;59(4):813-823. [doi: 10.1016/j.socscimed.2003.11.027] [Medline: 15177837]
32. DAGitty. Direct acyclic graph (DAG) on the determinants of attendance at health checks URL: http://www.dagitty.net [accessed 2018-09-27] [WebCite Cache ID 6zFVbmCxc]
33. Kass G. An Exploratory Technique for Investigating Large Quantities of Categorical Data. Applied Statistics 1980;29(2):119-127. [doi: 10.2307/2986296]
34. van Gelder MM, Bretveld R, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol 2010 Dec 01;172(11):1292-1298. [doi: 10.1093/aje/kwq291] [Medline: 20880962]
35. Christensen AI, Ekholm O, Davidsen M, Juel K. Viden. Copenhagen: Statens Institut for Folkesundhed; 2012. Health and morbidity in Denmark 2010 - and developments since 1987 URL: https://viden.sl.dk/artikler/voksne/socialt-udsatte/sundhed-og-sygelighed-2010/
36. Laws RA, Fanaian M, Jayasinghe UW, McKenzie S, Passey M, Davies GP, et al. Factors influencing participation in a vascular disease prevention lifestyle program among participants in a cluster randomized trial. BMC Health Serv Res 2013 May 31;13:201 [FREE Full text] [doi: 10.1186/1472-6963-13-201] [Medline: 23725521]
37. Bruun Larsen L, Soendergaard J, Halling A, Thilsing T, Thomsen J. A novel approach to population-based risk stratification, comprising individualized lifestyle intervention in Danish general practice to prevent chronic diseases: Results from a feasibility study. Health Informatics J 2017 Dec;23(4):249-259. [doi: 10.1177/1460458216645149] [Medline: 27245672]
38. Bjerregaard A, Maindal H, Bruun N, Sandbæk A. Patterns of attendance to health checks in a municipality setting: the Danish 'Check Your Health Preventive Program'. Prev Med Rep 2017 Mar;5:175-182 [FREE Full text] [doi: 10.1016/j.pmedr.2016.12.011] [Medline: 28050340]

39. Larsen LB, Sandbaek A, Thomsen JL, Bjerregaard A. Uptake of health checks by residents from the Danish social housing sector - a register-based cross-sectional study of patient characteristics in the 'Your Life - Your Health' program. BMC Public Health 2018 May 02;18(1):585 [FREE Full text] [doi: 10.1186/s12889-018-5506-6] [Medline: 29720162]
40. Norman P. Applying the health belief model to the prediction of attendance at health checks in general practice. Br J Clin Psychol 1995 Sep;34 (Pt 3):461-470. [Medline: 8845784]
41. Bender A, Jørgensen T, Helbech B, Linneberg A, Pisinger C. Socioeconomic position and participation in baseline and follow-up visits: the Inter99 study. Eur J Prev Cardiol 2014 Jul;21(7):899-905. [doi: 10.1177/2047487312472076] [Medline: 23233551]
42. Brunner-Ziegler S, Rieder A, Stein KV, Koppensteiner R, Hoffmann K, Dorner TE. Predictors of participation in preventive health examinations in Austria. BMC Public Health 2013 Dec 05;13:1138 [FREE Full text] [doi: 10.1186/1471-2458-13-1138] [Medline: 24308610]
43. Wilhelmsen L, Ljungberg S, Wedel H, Werkö L. A comparison between participants and non-participants in a primary preventive trial. J Chronic Dis 1976 May;29(5):331-339. [Medline: 939796]
44. Lakerveld J, Ijzelenberg W, van Tulder MW, Hellemans IM, Rauwerda JA, van Rossum AC, et al. Motives for (not) participating in a lifestyle intervention trial. BMC Med Res Methodol 2008 Apr 10;8:17 [FREE Full text] [doi: 10.1186/1471-2288-8-17] [Medline: 18402683]
45. Jacobsen B, Thelle D. The Tromsø Heart Study: responders and non-responders to a health questionnaire, do they differ? Scand J Soc Med 1988;16(2):101-104. [Medline: 3387934]
46. Thorogood M, Coulter A, Jones L, Yudkin P, Muir J, Mant D. Factors affecting response to an invitation to attend for a health check. J Epidemiol Community Health 1993 Jun;47(3):224-228 [FREE Full text] [Medline: 8350036]
47. Labeit A, Peinemann F, Baker R. Utilisation of preventative health check-ups in the UK: findings from individual-level repeated cross-sectional data from 1992 to 2008. BMJ Open 2013 Dec 23;3(12):e003387 [FREE Full text] [doi: 10.1136/bmjopen-2013-003387] [Medline: 24366576]
48. Koopmans B, Nielen MM, Schellevis FG, Korevaar JC. Non-participation in population-based disease prevention programs in general practice. BMC Public Health 2012 Oct 09;12:856 [FREE Full text] [doi: 10.1186/1471-2458-12-856] [Medline: 23046688]
49. Gruber J, Arnold B, Reygadas F, Hubbard A, Colford J. Estimation of treatment efficacy with complier average causal effects (CACE) in a randomized stepped wedge trial. Am J Epidemiol 2014;179(9):1134-1142.
50. Meader N, King K, Moe-Byrne T, Wright K, Graham H, Petticrew M, et al. A systematic review on the clustering and co-occurrence of multiple risk behaviours. BMC Public Health 2016 Dec 29;16:657 [FREE Full text] [doi: 10.1186/s12889-016-3373-6] [Medline: 27473458]
51. Robson J, Dostal I, Madurasinghe V, Sheikh A, Hull S, Boomla K, et al. NHS Health Check comorbidity and management: an observational matched study in primary care. Br J Gen Pract 2017 Feb;67(655):e86-e93 [FREE Full text] [doi: 10.3399/bjgp16X688837] [Medline: 27993901]
52. Woringer M, Cecil E, Watt H, Chang K, Hamid F, Khunti K, et al. Evaluation of community provision of a preventive cardiovascular programme - the National Health Service Health Check in reaching the under-served groups by primary care in England: cross sectional observational study. BMC Health Serv Res 2017 Dec 14;17(1):405 [FREE Full text] [doi: 10.1186/s12913-017-2346-5] [Medline: 28615019]
53. Roberts DJ, de Souza VC. A venue-based analysis of the reach of a targeted outreach service to deliver opportunistic community NHS Health Checks to 'hard-to-reach' groups. Public Health 2016 Aug;137:176-181. [doi: 10.1016/j.puhe.2016.03.004] [Medline: 27062066]
54. Thygesen L, Daasnes C, Thaulow I, Brønnum-Hansen H. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. Scand J Public Health 2011 Jul;39(7 Suppl):12-16. [doi: 10.1177/1403494811399956] [Medline: 21898916]

Abbreviations

ATC: anatomical therapeutic chemical
CHAID: chi-square automatic interaction detection
COPD: chronic obstructive pulmonary disease
CVD: cardiovascular disease
DAG: direct acyclic graph
EPR: electronic patient record
FEV1: forced expiratory volume in one second
FVC: forced vital capacity
GP: general practitioner
ICD-10: International Classification of Diseases 10th Edition
IRR: incidence rate ratio
SES: socioeconomic status

T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 22.07.18; peer-reviewed by J Ebert, J Apolinário-Hagen; comments to author 19.09.18; revised version received 27.09.18; accepted 29.09.18; published 21.01.19.

Please cite as:
Larsen LB, Sondergaard J, Thomsen JL, Halling A, Sønderlund AL, Christensen JR, Thilsing T
Digital Recruitment and Acceptance of a Stepwise Model to Prevent Chronic Disease in the Danish Primary Care Sector: Cross-Sectional Study
J Med Internet Res 2019;21(1):e11658
URL: http://www.jmir.org/2019/1/e11658/
doi:10.2196/11658
PMID:30664466

Original Paper

Expertise Modulates Students' Perception of Pain From a Self-Perspective: Quasi-Experimental Study

Sareh Said Yekta-Michael[1], DDS, PhD; André Schüppen[2], Dipl-Inform; Arnim Johannes Gaebler[3], Dr med; Jens Ellrich[4], MD, PhD; Jan Willem Koten[5], PhD

[1]Department of Conservative Dentistry, Periodontology and Preventive Dentistry, RWTH Aachen University, Aachen, Germany

[2]Interdisciplinary Center for Clinical Research, RWTH Aachen University, Aachen, Germany

[3]Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, RWTH Aachen, Aachen, Germany

[4]Medical Faculty, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen-Nuremberg, Germany

[5]Institute of Psychology, University of Graz, Graz, Austria

Corresponding Author:
Sareh Said Yekta-Michael, DDS, PhD
Department of Conservative Dentistry, Periodontology and Preventive Dentistry
RWTH Aachen University
Pauwelsstrasse 30
Aachen, 52074
Germany
Phone: 49 02418088118
Email: samichael@ukaachen.de

Abstract

Background: Perception of stimuli presented in a virtual dentistry environment affects regions of the brain that are related to pain perception.

Objective: We investigated whether neural correlates of virtual pain perception are affected by education in dentistry.

Methods: In this functional magnetic resonance imaging study, a sample of 20 dental students and 20 age-matched controls viewed and listened to video clips presenting a dental treatment from the first - person perspective. An anxiety questionnaire was used to assess the level of dental anxiety. Neural correlates of pain perception were investigated through classic general linear model analysis and in-house classification methods.

Results: Dental students and naïve controls exhibited similar anxiety levels for invasive stimuli. Invasive dentistry scenes evoked a less affective component of pain in dental students compared with naïve controls (P<.001). Reduced affective pain perception went along with suppressed brain activity in pain matrix regions including the insula, anterior cingulate cortex, and basal ganglia. Furthermore, a substantial reduction of brain activity was observed in motor-related regions, particularly the supplementary motor area, premotor cortex, and basal ganglia. Within this context, a classifier analysis based on neural activity in the nucleus lentiformis could identify dental students and controls on the individual subject level in 85% of the cases (34 out of 40 participants, sensitivity=90%, specificity=80%).

Conclusions: Virtual dental treatment activates pain-related brain regions in controls. By contrast, dental students suppress affective and motor-related aspects of pain. We speculate that dental students learn to control motoric aspects of pain perception during their education because it is a prerequisite for the professional manual treatment of patients. We discuss that a specific set of learning mechanisms might affect perceived self-efficacy of dental students, which in turn might reduce their affective component of pain perception.

(J Med Internet Res 2019;21(1):e10885) doi:10.2196/10885

KEYWORDS

medical education; virtual reality; questionnaires; physician

Introduction

Background

The quality of dentistry relies on the extent to which dentists can accurately plan, perform, and perceive motor actions. Planning performance and perception of potentially hurtful motor actions affect brain regions that are related to working memory and pain perception [1,2].

Pain is a variegated feeling, and it has a multidimensional nature with sensory-discriminative, a ective-motivational, motoric, and cognitive components [2-7]. Currently, neural correlates of orofacial pain are investigated through functional magnetic resonance imaging (fMRI) [8,9]. The affective component of pain is a subjective feeling that is often measured with standardized psychological scales that report the unpleasantness of stimuli. It is not only perceived when subjects are treated with physically painful stimuli but also when individuals are confronted with psychologically painful images or videos. However, the ability to perceive the affective component of pain is modulated by the professional training of a subject. Viewing hurting scenes induces affective component of pain in controls, whereas this is less the case for medical doctors [10,11]. fMRI studies show that medical doctors suppress the pain matrix when they view painful actions executed on others, whereas this is not the case for controls [10]. The pain matrix includes the thalamus, SI and SII, insula, as well as anterior cingulate cortex (ACC) [12-14]. In addition, cortical and subcortical motor regions including the basal ganglia are directly responsive to noxious stimuli [13,15]. Some of these cortical and subcortical motor regions show a nociceptive somatotopic organization [16].

Further, electroencephalogram studies suggest that early and late signal components show dissimilar behavior when controls view painful scenes, whereas this is not the case for doctors [11].

In addition, imaging studies showed that doctors and controls exhibit similar empathy, emotional contagion, and interpersonal reactivity scores [10,11].

Statistically valid differences between the 2 groups are exclusively found for scales that measure affective and sensory aspects of pain. From these neuroimaging studies, we have to conclude that affective and sensory aspects of pain for others may be modulated by education, whereas empathy for others as measured with several accepted scales is not. This leads to the somewhat counterintuitive conclusion that doctors maintain empathy for others, whereas suppressing neural correlates of pain for others [10,11,17]. We do not think that more research in the field of empathy is needed simply because previous empathy measurers failed to show differences between doctors and controls.

However, observed differences in affective pain perception for others might be linked to how doctors perceive pain from a self-perspective. Furthermore, one might speculate that differences in pain perception originate from medical training and occur in medical school.

Objectives

In this paper, we investigate the differences of brain activations during virtual dental treatment in dentistry students and controls. We choose to investigate this through classic general linear model (GLM) analysis and in-house classification methods. Classification methods may have some advantages over classic analysis. First, one can specifically test whether hypothetical brain regions identified in other studies may in fact contribute to the identification of different neuropsychological states. Within this context, we focused on previously reported coordinates that were related to different components of pain perception [13]. Second, although classic fMRI methods only focus on univariate measures of brain activity, classification methods can identify the multivariate interplay among brain regions. These multivariate aspects of brain activity may be better predictors of a certain neuropsychological state as univariate measures. Finally, classification methods may isolate brain regions that are essential for a specific function.

We hypothesize that watching dentistry scenes from a self-perspective may induce brain activity in (affective and motoric aspects) pain-related regions in controls, whereas this is less the case for dental students. We expect that dental students learn to control the motoric aspects of pain during their education because it is a prerequisite for manual treatment [1].

Methods

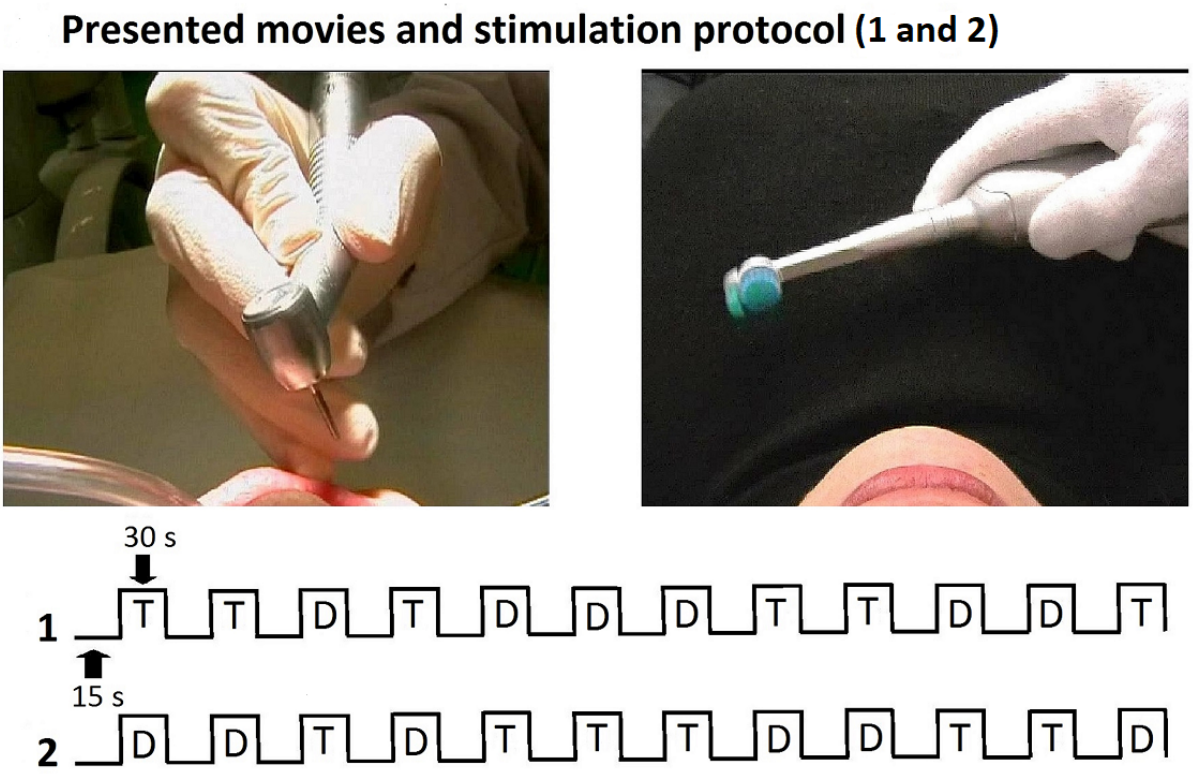
A total of 40 healthy (20 dental students, 20 controls), right-handed male volunteers (average age 28 years, SD 9) participated in this fMRI study [18] (Multimedia Appendix 1).

All subjects had been dentally treated in the recent past and had participated in former fMRI experiments. Controls were selected in the course of dental routine checkups in the Department of Conservative Dentistry (RWTH Aachen University). Dental students were selected 1 year before graduating. All volunteers gave their approval to the experimental conditions in written form.

Procedures were approved by the local ethics committee of University Hospital of Aachen, and all volunteers agreed to the World Medical Association Declaration of Helsinki (1964) and subsequent amendments.

The participants viewed and listened to drilling and toothbrush movies (Figure 1). The subjects were instructed to imagine the dental treatment from the perspective of a patient (first-person perspective). The sound of drilling and toothbrush movies was the same. The total duration of this procedure was 9 min.

Figure 1. Stimulation protocol. Drilling movies presented a medical glove–wearing hand with a dental handpiece drilling a tooth in the right lower jaw. Toothbrush movies displayed the same gloved-hand using an electric toothbrush. Every single movie was presented 12 times in counterbalanced order for 30 seconds and separated by 12 resting baseline conditions that lasted 15 seconds. Both movies were presented in a randomized fashion to the volunteers.



Anxiety Questionnaire

The Hierarchical Anxiety Questionnaire was used to assess the intensity of dental anxiety [19]. On the basis of an overall score ranging from 11 to 55, participants can be categorized into low anxious (<30), moderately anxious (31-38), and highly anxious (>38) groups.

Pain Perception Questionnaire

The Pain Perception Scale is a common standard instrument for the study of pain, allowing standardized and multifaceted quantification of pain experience. By default, it contains 19 sensory and 14 affective descriptions of pain. All the subjects were assigned to rank each description for drilling and tooth brushing on a 4-point scale immediately after the fMRI session [20].

Functional Magnetic Resonance Imaging Analysis

Scanning was performed by a Philips 3-Tesla magnetic resonance imaging *Model Achieva* (Philips Medical Systems). Axial slices were oriented toward the anterior-posterior commissure. A T2*-weighted echo planar imaging sequence was obtained for functional images: echo time 30 ms, repetition time 2800 ms, 32 interleaved slices (3.5 mm thick), flip angle of 90 degrees, field of view of 220 mm, voxel size of 3.75×3.75×3.5 mm, and 64×64 matrix. Functional data were imported into the Statistical Parametric Mapping (SPM) toolbox and coregistered with the high-resolution anatomical scan of the subject that was obtained in the same session. Preprocessing steps included the following: realignment, normalization to Montreal Neurological Institute (MNI) space, spatial smoothing (8 mm), and high-pass filtering (128 seconds). First-level beta weights were obtained by modeling the canonical hemodynamic response function within a GLM approach. Beta weights were used to contrast experimental conditions and experimental groups on a second level using a 2-sample *t* test. Furthermore, 2 contrasts were reported. First, we wanted to know if controls exhibit higher brain activity compared with dental students when the toothbrush condition was subtracted from the drilling condition ($drill_C$−$toothbrush_C$)−($drill_{DS}$−$toothbrush_{DS}$). The latter contrast was liberally masked with the contrast $drill_C$−$toothbrush_C$ to avoid spurious activations. Next, we wanted to know if dental students exhibit higher brain activity compared with controls when toothbrush was subtracted from drilling. The latter contrast was masked with the $drill_{DS}$−$toothbrush_{DS}$. We thresholded contrasts at P<.001. Subsequently, we performed a Monte Carlo–based cluster threshold estimation procedure to correct for multiple testing. Next, beta contrast weights (drill−toothbrush) were extracted from 40 pain relevant cortical systems reported in a meta-analysis [13]. Extracted beta weights were subjected to a support vector machine analysis.

Classifier Analysis

As information-based procedure to determine differences in brain activity between both groups on the individual subject level, we conducted a classifier analysis. Therefore, a modified support vector machine algorithm with a leave-one-out cross-validation was applied [21].

Region of Interest Definition and Feature Generation

Beta contrast weights (drilling−toothbrush) were extracted from pain-relevant brain regions reported in a meta-analysis [13]. For this purpose, structural scans were segmented into gray and white matter using the standard tools as available in the SPM software package. Regions of interest were defined, centering 4-mm diameter spheres on MNI coordinates of 40 pain-relevant brain regions reported in the abovementioned meta-analysis. Beta weights were extracted only from those voxels within the sphere which were found within the gray matter. This method avoids the extraction of spurious beta weights. Next, we averaged the beta weights per region. We subtracted beta weights of the toothbrush conditions from drill conditions for every region and subject. Functional masking of the beta weights was not needed because beta contrast weights of the drilling were positive and larger than the respective beta weights of the toothbrush conditions in all regions in all subjects.

Feature Selection

To consider only the most discriminating features for the classifier analysis, an information-based feature selection was applied. The discriminative power of a feature was defined as the absolute value of the Kendall tau correlation coefficient [22], which measured the correlation between a feature and the group indicator (−1 for controls, +1 for dental students). Thus, a positive correlation coefficient indicates that the feature (ie, the regional brain activity) increases in dental students compared with controls, whereas a negative correlation coefficient indicates that the feature decreases in the dental students compared with controls.

In each fold of the leave-one-out cross-validation, the features obtained from the n-1 remaining subjects were ranked according to their absolute value of the Kendall tau rank correlation coefficient and the feature, which exhibited highest relation to the group indicator, was selected.

Classifiers

The selected features were subjected to the classifier analysis applying support vector machines with a linear kernel [23]. Therefore, the support vector machine yielded a maximal-margin hyperplane in the feature space, which separated the groups in the respective training dataset. Classification was tested in the left-out sample. The limited number of subjects was lent to the leave-one-out cross-validation method to investigate the generalizability of the classification results. Importantly, this cross-validation encompassed the feature selection as well as the classifier. Accuracy (percentage of all participants detected correctly), sensitivity (percentage of dental students detected correctly), and specificity (percentage of controls detected correctly) quantified classification performance.

Results

After the fMRI session, all controls and dental students declared that they could imagine being treated by a dental drill from a first-person perspective.

Anxiety Questionnaire

From the 40 individuals under study, 12 controls and 10 dental students were categorized as low anxious, 8 controls and 8 dental students were categorized as moderately anxious, and 2 dental students were categorized as highly anxious (average controls 29, average dental students 30).

Pain Perception Questionnaire

We chose to analyze data with a very conservative approach. For every condition, 33 *t* tests were executed according to the number of items in the questionnaire. Next, the critical Bonferroni threshold .05/33 was estimated. The summary statistic based on 2 sample *t* tests is visualized (Figure 2). No significant differences between dental students and controls were observed for toothbrush, but large differences were observed for drilling. For drilling, significant differences between dental students and controls were observed for affective pain scales but not for sensory pain scales.

Dental students showed significantly lower pain than controls for 4 affective items, namely agonizing, dreadful, horrible, and enervating ($P<.001$, Bonferroni).

Neural aspects of pain perception were not higher in controls compared with dental students when toothbrush was shown. Moreover, neural aspects of pain perception were not higher in dental students compared with controls when drilling was compared with toothbrush (empty contrast). By contrast, neural aspects of pain perception were higher in controls than in dental students when drilling was compared with toothbrush (Figures 3 and 4; Multimedia Appendix 2).

Figure 2. Anxiety Questionnaire. Comparison controls versus dental students (mean ± standard error of the mean, each group n=20; 1: not appropriate, 2: somewhat appropriate, 3: largely appropriate, 4: fully appropriate). Asterisks indicate significant differences between the ratings for dental students and controls for the drilling movies (one asterisk denotes *P*<.05, double asterisks denote *P*<.01, triple asterisks denote *P*<.001, *t* test). Red Asterisks: Bonferroni threshold.

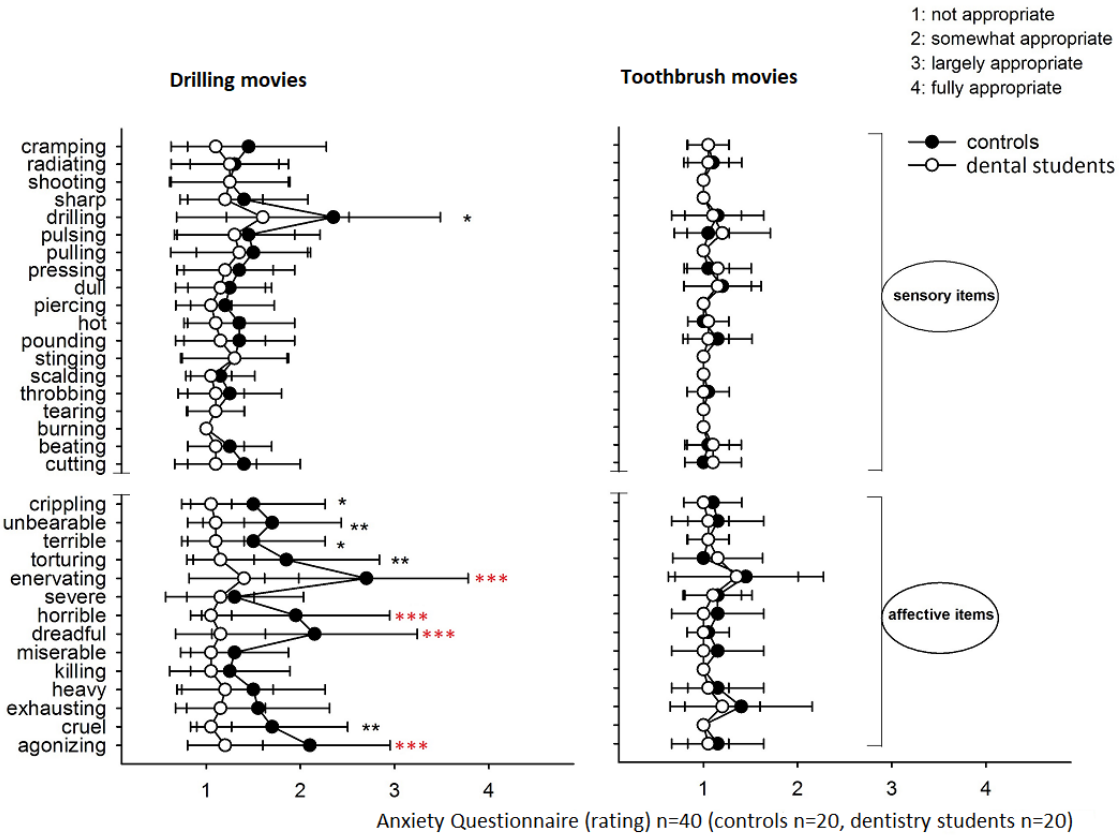


Figure 3. Brain activity. Controls exhibit higher brain activity compared with dental students when the toothbrush conditions were subtracted from the drilling conditions (drillC-toothbrushC)-(drillDS-toothbrushDS). The latter contrast was liberally masked with (drillNC-toothbrushC). Contrasts were thresholded at P<.001 with k=156.

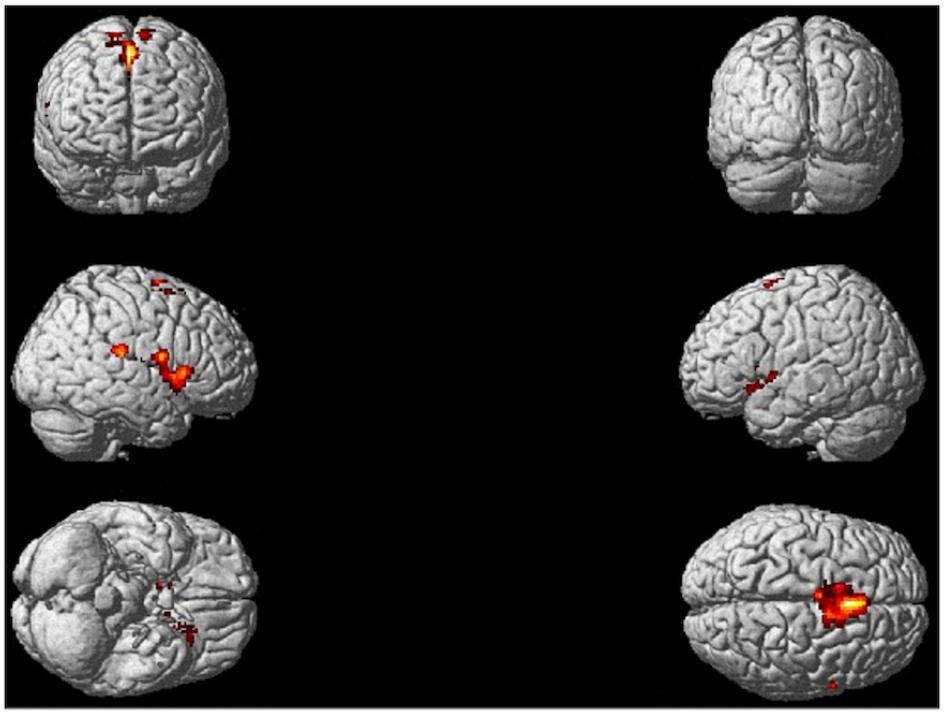
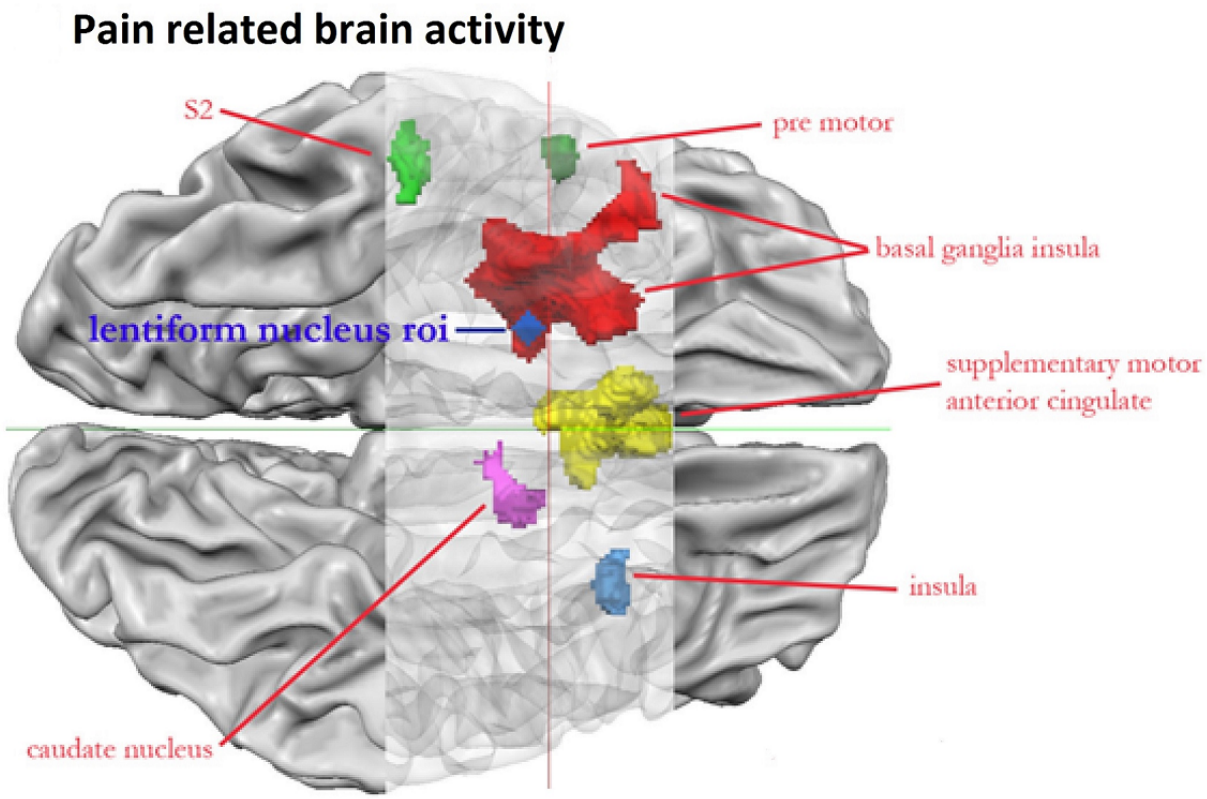


Figure 4. Region of interest (roi). Controls exhibit higher brain activity compared with dental students when the toothbrush conditions were subtracted from the drilling conditions (drillC-toothbrushC)-(drillDS-toothbrushDS). The latter contrast was liberally masked with (drillC-toothbrushC). Contrasts were thresholded at *P*<.0005 with k=125.In addition, we visualize the roi that leads to best classification results.



Functional Magnetic Resonance Imaging Analysis

Results (*P*<.001 using a conservative cluster threshold of *k*=159; Figure 3) revealed that brain activations were organized around sensory-motor and limbic-affective systems. The sensory motor system included larger parts of the supplementary motor area as well as SII. Furthermore, limbic-affective structures included the thalamus basal ganglia as well as larger parts of the posterior and frontal insula and ACC. It is clear that some structures can be part of more systems. For instance, the basal ganglia are part of the motor loop. We used the standard SPM preprocessing pipeline that employs rather crude smoothing kernels (8 mm) and rough brain alignment methods. This may lead to spurious activations even when conservative cluster thresholds are used. In a next step, we decided to threshold our image at .0005 in combination with a less conservative cluster threshold (*k*=125). Some of the previously discussed activations disappeared. Unfortunately, the SPM brain rendering does not fully inform about subcortical and insular activations. For this reason, a ventral view of the activation is presented in Figure 4 using the brain voyager software.

In dental students, brain activity in a web of brain regions known as the pain matrix was suppressed. Remarkably enough, the difference between dental students and controls was not found in regions that are of core importance for the processing of the sensorial aspects of pain such as the SI. However, dental students showed a marked suppression of brain activity in regions related to the affective aspects of pain including the bilateral insula and bilateral ACC. In addition, motor-related aspects of pain perception were reduced in dental students. The latter included the premotor cortex as well as the nuclei of the basal ganglia and caudate nucleus (Figures 3 and 4). We want to stress that we could not identify brain regions that were activated in dental students but not in controls.

Classifier Analysis

In the classifier analysis, we were able to classify 85% (34 participants of 40) of the participants on the basis of neural activity found in the most discriminative region in each fold of the cross-validation (sensitivity=90.0%, specificity=80.0%). It turned out that in each fold of the leave-one-out cross-validation, the neural activity in the right nucleus lentiformis (dark blue region in Figure 4) was selected as the most discriminative feature. Adding more regions to the classifier did not result in better classification performance.

Discussion

Principal Findings

In this study, we showed that dentistry students suppress pain-linked brain activity when brought into a virtual dental-treatment environment, whereas this is less the case for controls. In addition, we could identify dental students and controls on the basis of brain activity located in the lentiform nucleus.

In this study, affective items of Anxiety questionnaire clearly indicated a difference between controls and dental students.

Controls perceived drilling as more unpleasant when compared with dental students. By contrast, no affective pain perception differences were observed for the toothbrush. In addition, both groups showed similar scores with regard to sensory aspects of pain regardless of the condition type under study. This indicates that dental students experience less affective component of pain than controls when viewing invasive dentistry scenes. The reduced pain perception of dental students correlated with the reduced brain activity of pain matrix regions. The latter included regions related to the somatosensory system as well as regions that have been related to affective aspects of pain perception including parts of the insula [24] and ACC [12,25]. We also observed reduced activity of dental students in subcortical and cortical motor regions including precentral gyrus (Broadman area 6) and the basal ganglia, particularly the lentiform nucleus. The observed differences in motor regions suggest that dental students suppress motoric components of pain when confronted with invasive stimuli. In short, most of our hypotheses were confirmed.

A difference between previous studies and this study is that we assessed affective pain perception from a self-perspective, whereas Cheng and Decety investigated affective pain from another perspective. Despite the obvious difference in perspective, we reproduce some important findings of Chen et al. In both studies, controls exhibited increased activity in ACC and the supplementary motor area when confronted with invasive stimuli [26]. However, Cheng observed that experts activate a frontal parietal system when confronted with invasive stimuli. They speculated that these increases in brain activation were linked to emotion regulation and theory of mind. We did not observe increased brain activity in dental students. It may be possible that observed differences between the 2 studies are because of the self-perspective versus other perspective. However, one should be careful with these kinds of speculations. In fact, the study of Cheng lacked a control condition from a self-perspective, whereas this study lacks a control condition from the other perspective.

This study also shows the benefits of classification approaches. We expected that a larger number of regions was needed to obtain sufficient classification rates [21,27]. However, in fact, only 1 region, namely the lentiform nucleus, was needed to classify controls and dental students on the basis of brain activity. This suggests that the latter region is of core importance in pain perception. Although conventional GLM analysis may be used to trace differences between groups, it is maybe not the ideal method to isolate brain regions that are of core importance for a specific function. Hence, classification methods may be a useful complement to conventional GLM methods.

Our findings suggest that affective and motoric components of pain suppression of dental students might possibly originate from dental school training. Recently, it has been suggested that distinct learning mechanisms affect pain expectancy, which in turn affects pain perception [28]. According to Peerdeman, pain perception can be affected by cognitive instructions, observational learning, and operant conditioning. The 3 learning mechanisms mentioned may modulate pain expectancy in medical students. Pain expectancy has been linked to activity in the insula [29] and basal ganglia [30]. Our results suggest that missing activity in the insula and basal ganglia in dental students reflects lower anticipation of pain. Furthermore, Peerdeman argues that the aforementioned learning mechanisms may affect self-efficacy expectancy. This is defined as the extent to which people can voluntary control aspects of pain. Neural correlates of self-efficacy have been linked to the nucleus lentiformis [31]. As demonstrated in this study, brain activity of this region predicts whether an individual belongs to the dental students or controls. As mentioned above, we speculate that maintaining motor control in the face of pain is an important ingredient of medical education. A previous study investigated the cognitive aspects of motor actions in a virtual dentistry environment in a sample of dentistry students [1]. However, the quality of the dentistry treatment may not only depend on the cognitive aspects of motor actions but also on emotional aspects of motor actions. The latter was the main object of our virtual-reality study, and we therefore think that it completes the previous motor study.

Limitations

As a limitation, the sample size of this study is 40 subjects. Although this is perfectly normal for an fMRI study, one might argue that the power to detect effects is not very large. We would like to argue that we replicate important findings of our colleagues.

There are some limitations in our experimental design. First, it might be better to study the effect of hurtful scenes from both a self- and another perspective. Second, it might be better to study dental students at the end and beginning of their studies using a within-subject design. Finally, one might investigate if individual differences in pain perception of dental students are linked to their clinical performance [17]. In previous studies, subjects viewed hurtful actions inflicted on others, whereas in this study subjects viewed hurtful actions inflicted on the subject itself [8-11]. It is indeed possible that the expertise obtained during medical school leads to a suppression of pain from a self-perspective, which in turn might affect pain perception from another perspective. However, in principle, an opposite mechanism is possible, namely reduced pain perception for others leads to reduced pain perception for oneself. Unfortunately, it is not possible to identify the exact causal relation in this study.

Conclusions

We conclude that dental students suppress affective and motor-related aspects of pain. The cognitive mechanisms that modulate pain expectancy and pain perception are poorly understood and deserve further investigation. A candidate mechanism is self-efficacy that may be linked to brain activity in the nucleus lentiformis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Consolidated Standards of Reporting Trials (CONSORT)-flow diagram.

[PDF File (Adobe PDF File), 83KB - jmir_v21i1e10885_app1.pdf]

Multimedia Appendix 2

The statistics for brain coordinates.

[PDF File (Adobe PDF File), 32KB - jmir_v21i1e10885_app2.pdf]

References

1. Perry S, Bridges SM, Zhu F, Leung WK, Burrow MF, Poolton J, et al. Getting to the root of fine motor skill performance in dentistry: brain activity during dental tasks in a virtual reality haptic simulation. J Med Internet Res 2017 Dec 12;19(12):e371 [FREE Full text] [doi: 10.2196/jmir.8046] [Medline: 29233801]
2. Lin CS, Wu SY, Yi CA. Association between anxiety and pain in dental treatment: a systematic review and meta-analysis. J Dent Res 2017 Dec;96(2):153-162. [doi: 10.1177/0022034516678168] [Medline: 28106507]
3. Said Yekta S, Vohn R, Ellrich J. Cerebral activations resulting from virtual dental treatment. Eur J Oral Sci 2009 Dec;117(6):711-719. [doi: 10.1111/j.1600-0722.2009.00689.x] [Medline: 20121935]
4. Meier ML, Widmayer S, Abazi J, Brügger M, Lukic N, Lüchinger R, et al. The human brain response to dental pain relief. J Dent Res 2015 May;94(5):690-696. [doi: 10.1177/0022034515572022] [Medline: 25691071]
5. Lin CS, Niddam DM, Hsu ML, Hsieh JC. Pain catastrophizing is associated with dental pain in a stressful context. J Dent Res 2013 Feb;92(2):130-135 [FREE Full text] [doi: 10.1177/0022034512467804] [Medline: 23232145]
6. Mathur VA, Khan SA, Keaser ML, Hubbard CS, Goyal M, Seminowicz DA. Altered cognition-related brain activity and interactions with acute pain in migraine. Neuroimage Clin 2015;7:347-358. [doi: 10.1016/j.nicl.2015.01.003] [Medline: 25610798]
7. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital support interventions for the self-management of low back pain: a systematic review. J Med Internet Res 2017 May 21;19(5):e179 [FREE Full text] [doi: 10.2196/jmir.7290] [Medline: 28550009]
8. Chantaracherd P, John MT, Hodges JS, Schiffman EL. Temporomandibular joint disorders' impact on pain, function, and disability. J Dent Res 2015 Mar;94(3 Suppl):79S-86S. [doi: 10.1177/0022034514565793] [Medline: 25572112]
9. Schiffman EL, Ahmad M, Hollender L, Kartha K, Ohrbach R, Truelove EL, et al. Longitudinal stability of common TMJ structural disorders. J Dent Res 2017 Mar;96(3):270-276 [FREE Full text] [doi: 10.1177/0022034516679396] [Medline: 27856966]
10. Cheng Y, Lin CP, Liu HL, Hsu YY, Lim KE, Hung D, et al. Expertise modulates the perception of pain in others. Curr Biol 2007 Oct 09;17(19):1708-1713 [FREE Full text] [doi: 10.1016/j.cub.2007.09.020] [Medline: 17900903]
11. Decety J, Yang C, Cheng Y. Physicians down-regulate their pain empathy response: an event-related brain potential study. Neuroimage 2010 May 01;50(4):1676-1682. [doi: 10.1016/j.neuroimage.2010.01.025] [Medline: 20080194]
12. Apkarian AV, Bushnell MC, Treede RD, Zubieta JK. Human brain mechanisms of pain perception and regulation in health and disease. Eur J Pain 2005 Aug;9(4):463-484. [doi: 10.1016/j.ejpain.2004.11.001] [Medline: 15979027]
13. Duerden EG, Albanese MC. Localization of pain-related brain activation: a meta-analysis of neuroimaging data. Hum Brain Mapp 2013 Jan;34(1):109-149. [doi: 10.1002/hbm.21416] [Medline: 22131304]
14. Iadarola MJ, Max MB, Berman KF, Byas-Smith MG, Coghill RC, Gracely RH, et al. Unilateral decrease in thalamic activity observed with positron emission tomography in patients with chronic neuropathic pain. Pain 1995 Oct;63(1):55-64. [Medline: 8577491]
15. Chudler EH, Dong WK. The role of the basal ganglia in nociception and pain. Pain 1995 Jan;60(1):3-38. [doi: 10.1016/0304-3959(94)00172-B] [Medline: 7715939]
16. Bingel U, Gläscher J, Weiller C, Büchel C. Somatotopic representation of nociceptive information in the putamen: an event-related fMRI study. Cereb Cortex 2004 Dec;14(12):1340-1345. [doi: 10.1093/cercor/bhh094] [Medline: 15217895]
17. Jensen KB, Petrovic P, Kerr CE, Kirsch I, Raicek J, Cheetham A, et al. Sharing pain and relief: neural correlates of physicians during treatment of patients. Mol Psychiatry 2014 Mar;19(3):392-398 [FREE Full text] [doi: 10.1038/mp.2012.195] [Medline: 23358155]
18. 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] [Medline: 22209829]

19. Enkling N, Marwinski G, Jöhren P. Dental anxiety in a representative sample of residents of a large German city. Clin Oral Investig 2006 Mar;10(1):84-91. [doi: 10.1007/s00784-006-0035-6] [Medline: 16477408]
20. Geissner E. [The Pain Perception Scale--a differentiated and change-sensitive scale for assessing chronic and acute pain]. Rehabilitation (Stuttg) 1995 Nov;34(4):XXXV-XLIII. [Medline: 8570898]
21. Gaebler A, Mathiak K, Koten JJ, König AA, Koush Y, Weyer D, et al. Auditory mismatch impairments are characterized by core neural dysfunctions in schizophrenia. Brain 2015 May;138(Pt 5):1410-1423 [FREE Full text] [doi: 10.1093/brain/awv049] [Medline: 25743635]
22. Vexler A, Chen X, Hutson AD. Dependence and independence: Structure and inference. Stat Methods Med Res 2017 Oct;26(5):2114-2132. [doi: 10.1177/0962280215594198] [Medline: 26229085]
23. Misaki M, Kim Y, Bandettini PA, Kriegeskorte N. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. Neuroimage 2010 Oct 15;53(1):103-118 [FREE Full text] [doi: 10.1016/j.neuroimage.2010.05.051] [Medline: 20580933]
24. Mazzola V, Latorre V, Petito A, Gentili N, Fazio L, Popolizio T, et al. Affective response to a loved one's pain: insula activity as a function of individual differences. PLoS One 2010 Dec 16;5(12):e15268 [FREE Full text] [doi: 10.1371/journal.pone.0015268] [Medline: 21179564]
25. Rotge JY, Lemogne C, Hinfray S, Huguet P, Grynszpan O, Tartour E, et al. A meta-analysis of the anterior cingulate contribution to social pain. Soc Cogn Affect Neurosci 2015 Jan;10(1):19-27 [FREE Full text] [doi: 10.1093/scan/nsu110] [Medline: 25140048]
26. Chavalarias D, Wallach JD, Li AH, Ioannidis JP. Evolution of Reporting P Values in the Biomedical Literature, 1990-2015. J Am Med Assoc 2016 Mar 15;315(11):1141-1148. [doi: 10.1001/jama.2016.1952] [Medline: 26978209]
27. Dinkel P, Willmes K, Krinzinger H, Konrad K, Koten Jr JW. Diagnosing developmental dyscalculia on the basis of reliable single case FMRI methods: promises and limitations. PLoS One 2013 Dec 09;8(12):e83722 [FREE Full text] [doi: 10.1371/journal.pone.0083722] [Medline: 24349547]
28. Peerdeman K, van Laarhoven AI, Peters M, Evers A. An integrative review of the influence of expectancies on pain. Front Psychol 2016 Aug 23;7:1270 [FREE Full text] [doi: 10.3389/fpsyg.2016.01270] [Medline: 27602013]
29. Sawamoto N, Honda M, Okada T, Hanakawa T, Kanda M, Fukuyama H, et al. Expectation of pain enhances responses to nonpainful somatosensory stimulation in the anterior cingulate cortex and parietal operculum/posterior insula: an event-related functional magnetic resonance imaging study. J Neurosci 2000 Oct 01;20(19):7438-7445 [FREE Full text] [doi: 10.1523/JNEUROSCI.20-19-07438.2000] [Medline: 11007903]
30. Keltner JR, Furst A, Fan C, Redfern R, Inglis B, Fields HL. Isolating the modulatory effect of expectation on pain transmission: a functional magnetic resonance imaging study. J Neurosci 2006 Apr 19;26(16):4437-4443 [FREE Full text] [doi: 10.1523/JNEUROSCI.4463-05.2006] [Medline: 16624963]
31. Kong J, Kaptchuk TJ, Polich G, Kirsch I, Vangel M, Zyloney C, et al. An fMRI study on the interaction and dissociation between expectation of pain relief and acupuncture treatment. Neuroimage 2009 Sep;47(3):1066-1076 [FREE Full text] [doi: 10.1016/j.neuroimage.2009.05.087] [Medline: 19501656]

Abbreviations

ACC: anterior cingulate cortex
fMRI: functional magnetic resonance imaging
GLM: general linear model
MNI: Montreal Neurological Institute
SPM: Statistical Parametric Mapping

Edited by G Eysenbach; submitted 16.05.18; peer-reviewed by S Edirippulige; comments to author 09.10.18; revised version received 08.11.18; accepted 12.12.18; published 23.01.19.

Please cite as:
Said Yekta-Michael S, Schüppen A, Gaebler AJ, Ellrich J, Koten JW
Expertise Modulates Students' Perception of Pain From a Self-Perspective: Quasi-Experimental Study
J Med Internet Res 2019;21(1):e10885
URL: http://www.jmir.org/2019/1/e10885/
doi:10.2196/10885
PMID:30674449

Original Paper

Identifying Common Methods Used by Drug Interaction Experts for Finding Evidence About Potential Drug-Drug Interactions: Web-Based Survey

Amy J Grizzle[1], PharmD; John Horn[2], PharmD; Carol Collins[2], MD; Jodi Schneider[3], PhD; Daniel C Malone[4], PhD; Britney Stottlemyer[5]; Richard David Boyce[5], PhD

[1]Center for Health Outcomes & PharmacoEconomic Research, College of Pharmacy, University of Arizona, Tucson, AZ, United States

[2]School of Pharmacy, University of Washington, Seattle, WA, United States

[3]School of Information Sciences, University of Illinois at Urbana-Champaign, Champaign, IL, United States

[4]Center for Health Outcomes & PharmacoEconomic Research, College of Pharmacy Department of Pharmacy Practice and Science, University of Arizona, Tucson, AZ, United States

[5]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Amy J Grizzle, PharmD
Center for Health Outcomes & PharmacoEconomic Research
College of Pharmacy
University of Arizona
PO Box 210202
1295 N Martin Ave
Tucson, AZ, 85721-0202
United States
Phone: 1 520 626 4721
Email: grizzle@pharmacy.arizona.edu

Abstract

Background: Preventing drug interactions is an important goal to maximize patient benefit from medications. Summarizing potential drug-drug interactions (PDDIs) for clinical decision support is challenging, and there is no single repository for PDDI evidence. Additionally, inconsistencies across compendia and other sources have been well documented. Standard search strategies for complete and current evidence about PDDIs have not heretofore been developed or validated.

Objective: This study aimed to identify common methods for conducting PDDI literature searches used by experts who routinely evaluate such evidence.

Methods: We invited a convenience sample of 70 drug information experts, including compendia editors, knowledge-base vendors, and clinicians, via emails to complete a survey on identifying PDDI evidence. We created a Web-based survey that included questions regarding the (1) development and conduct of searches; (2) resources used, for example, databases, compendia, search engines, etc; (3) types of keywords used to search for the specific PDDI information; (4) study types included and excluded in searches; and (5) search terms used. Search strategy questions focused on 6 topics of the PDDI information—(1) that a PDDI exists; (2) seriousness; (3) clinical consequences; (4) management options; (5) mechanism; and (6) health outcomes.

Results: Twenty participants (response rate, 20/70, 29%) completed the survey. The majority (17/20, 85%) were drug information specialists, drug interaction researchers, compendia editors, or clinical pharmacists, with 60% (12/20) having >10 years' experience. Over half (11/20, 55%) worked for clinical solutions vendors or knowledge-base vendors. Most participants developed (18/20, 90%) and conducted (19/20, 95%) search strategies without librarian assistance. PubMed (20/20, 100%) and Google Scholar (11/20, 55%) were most commonly searched for papers, followed by Google Web Search (7/20, 35%) and EMBASE (3/20, 15%). No respondents reported using Scopus. A variety of subscription and open-access databases were used, most commonly Lexicomp (9/20, 45%), Micromedex (8/20, 40%), Drugs@FDA (17/20, 85%), and DailyMed (13/20, 65%). Facts and Comparisons was the most commonly used compendia (8/20, 40%). Across the 6 attributes of interest, generic drug name was the most common keyword used. Respondents reported using more types of keywords when searching to identify the existence of PDDIs and determine their mechanism than when searching for the other 4 attributes (seriousness, consequences, management, and health

outcomes). Regarding the types of evidence useful for evaluating a PDDI, clinical trials, case reports, and systematic reviews were considered relevant, while animal and in vitro data studies were not.

Conclusions: This study suggests that drug interaction experts use various keyword strategies and various database and Web resources depending on the PDDI evidence they are seeking. Greater automation and standardization across search strategies could improve one's ability to identify PDDI evidence. Hence, future research focused on enhancing the existing search tools and designing recommended standards is needed.

(J Med Internet Res 2019;21(1):e11182) doi:10.2196/11182

KEYWORDS

drug interactions; drug interaction experts; potential drug-drug interactions; surveys

Introduction

One source of preventable harm related to medications is exposure to drug combinations that are known to interact. A recent meta-analysis of studies found that 22.2% of adverse drug event-associated hospital admissions were attributable to drug-drug interactions (308 drug-drug interaction cases/1683 patients, interquartile range, 16.6%-36.0%) [1]. Ensuring that medication therapy occurs safely and to the maximum benefit for any given patient is of great interest to clinicians [2]. Although computer-generated drug interaction alerts have the potential to provide clinicians with useful decision support, the poor specificity of alerting systems overwhelms clinicians with information that is difficult to use [3]. These factors may contribute to the >90% override rate consistently reported for clinicians [4]. Moreover, variability across electronic prescribing and pharmacy drug interaction alerting software systems is well documented and leads to clinician frustration and dissatisfaction [5-9]. A 2017 study of 3 commercial knowledge bases found substantial variability in the numbers of alerts generated for contraindicated and major or severe potential drug-drug interactions (PDDIs), with 25, 84, and 145 alerts per 1000 prescriptions for the 3 systems [10].

In prior work, we described the workflow of individuals who maintained PDDI information resources used by clinicians [11]. We refer to these individuals as *compendium editors*. The workflow of compendium editors generally involves topic identification, evidence search, evidence synthesis, and generating recommendations [11]. The evidence search step is nontrivial because there is no single repository housing data on PDDIs. Rather, there exist a wide variety of sources ranging from drug product labeling, regulatory documents, indexed scientific literature, to various knowledge bases and websites. The variety of sources makes it difficult to locate and synthesize the PDDI information into summaries that can help clinicians ensure that patients receive safe medication therapies. In addition to compendium editors, other drug interaction experts evaluate the PDDI information in response to client or colleague requests, assigned work tasks, or the availability of new evidence.

In an effort toward recommending a comprehensive search strategy that could effectively be used for identifying the relevant PDDI evidence, we surveyed drug interaction experts to better understand the various ways that they conduct PDDI evidence searches. This study aims to gather information to assist in designing candidate standard search strategies.

A "search strategy" is a systematic plan for locating relevant sources of information about some topics. Studies in the biomedical literature reported search strategies for the retrieval of clinical studies [12], diagnostic accuracy studies [13], animal studies [14], and studies about adverse events [15] among others. Conversely, very little research has been done to identify an optimal search strategy for the capture of complete and accurate PDDI information. Furthermore, this study aims to identify common methods for conducting PDDI literature searches used by experts who routinely evaluate such evidence.

Methods

To assist in the survey development, we examined whether similar research has been published. We conducted the following PubMed query in February 2018 and screened the results for relevant studies:

> "drug interactions"[MeSH Terms] AND "information storage and retrieval"[MeSH Terms] AND ("humans"[MeSH Terms] AND English[lang])

Of 460 results, 12 studies (2.6%) focused on data mining to identify PDDIs within titles, abstracts, and papers [16-28]. Only 1 paper evaluated a standardized search for identifying the PDDI evidence for specific drug pairs; this study focused on only one aspect of a search strategy—the search terms used to query the indexed scientific literature [28].

For this study, we incorporated other relevant search strategy aspects including (1) the list of the kinds of information being sought (eg, full-text papers vs abstracts; topical review vs primary research studies; regulatory documents vs news papers, etc) and (2) the sources of information one plans to search (eg, indexed scientific literature, books, conference proceedings, regulatory websites, etc).

A multidisciplinary team of investigators identified relevant attributes of PDDI evidence and, then, designed a survey incorporating these attributes. The team consisted of experts in drug interactions, literature searches (ie, librarian), health services researchers, and biomedical informatics. A librarian provided input to ensure a more comprehensive list of data sources for PDDI searches. Refinements were made, and, then, the instrument was integrated into a Web-based survey using Qualtrics software (Multimedia Appendix 1).

The final survey consisted of 16 questions grouped into 6 areas as follows: (1) work setting, experience, and area of expertise; (2) how to develop and conduct literature searches; (3) resources

used, including subscription and open-access databases, compendia, Web-based tools, and search engines for indexed literature; (4) keywords used to search for the specific PDDI information; (5) study types included and excluded in searches for the specific PDDI information; and (6) search terms used. The survey provided optional responses for resources, keywords, and study types that were identified from the team's prior research on the information needs of professionals who search and synthesize PDDI evidence [11]. Questions about search strategies were organized into 6 PDDI topics as follows: (1) that a PDDI exists; (2) seriousness; (3) clinical consequences; (4) management options; (5) mechanism; and (6) health outcomes. These categories were developed depending on the range of PDDI topics important for clinical decision support published by Payne et al [29]. Two of the survey questions provided open-ended responses that participants could use to provide additional comments, as well as share search terms that they find useful.

A list of drug interaction experts was assembled starting with individuals our team has worked with on past PDDI projects, including the aforementioned information needs study [11] and a prior workgroup on PDDI evidence assessment [30]. Experts were contacted and asked to provide names and contact information of other colleagues who would be appropriate to send the survey. A convenience sample of 70 experts, including compendia editors, knowledge-base vendors, and clinicians, were invited via emails to participate. The invitation and the survey introduction included a description and purpose of the survey, eligibility criteria, the estimated survey completion time of 15 minutes, and funding source. Those who agreed to participate were sent a link to the survey. Participants who completed the anonymous survey were compensated US $20 for their time. The University of Pittsburgh Institutional Review Board deemed this project exempt from human subject research requirements.

Results

Demographics

In this study, 20 of 70 invitees (response rate, 29%) completed the survey. The majority (17/20, 85%) of respondents were drug information specialists, drug interaction researchers, compendia editors, or clinical pharmacists. Over half of the participants (11/20, 55%) worked for clinical solutions vendors or knowledge-base vendors (Table 1). Furthermore, 60% (12/20) had >10 years of experience.

Developing and Conducting Searches

Most participants reported developing (18/20, 90%) and conducting (19/20, 95%) their own search strategies for PDDI evidence without assistance from a librarian. Few reported using Web-based search filters to assist developing (4/20, 20%) or conducting (3/20, 15%) searches. All 20 participants (20/20, 100%) reported using PubMed, and 11 (55%) included Google Scholar when searching abstracting services for published scientific papers (Figure 1). There were few reports of the use of EMBASE, Scopus, or Ovid MEDLINE.

Table 1. Participants' background.

Respondents' characteristics	n (%)
Professional title or job function	
Compendia editor	13 (65)
Drug information researcher	10 (50)
Drug information specialist	9 (45)
Clinical pharmacist	9 (45)
Systems analyst or content specialist	3 (15)
Pharmacy and therapeutics committee	2 (10)
Physician	1 (5)
Regulatory scientist	0 (0)
Other (Drug-drug interaction surveillance databases)	1 (5)
Work settings	
Clinical solutions vendor	7 (35)
Academic institution	6 (30)
Knowledge-base vendor	4 (20)
Drug information center	2 (10)
Hospital	1 (5)
Regulatory or government agency	0 (0)

Figure 1. The number of participants reporting using abstracting services to find indexed scientific literature for potential drug-drug interaction (PDDI) searches (N=20).

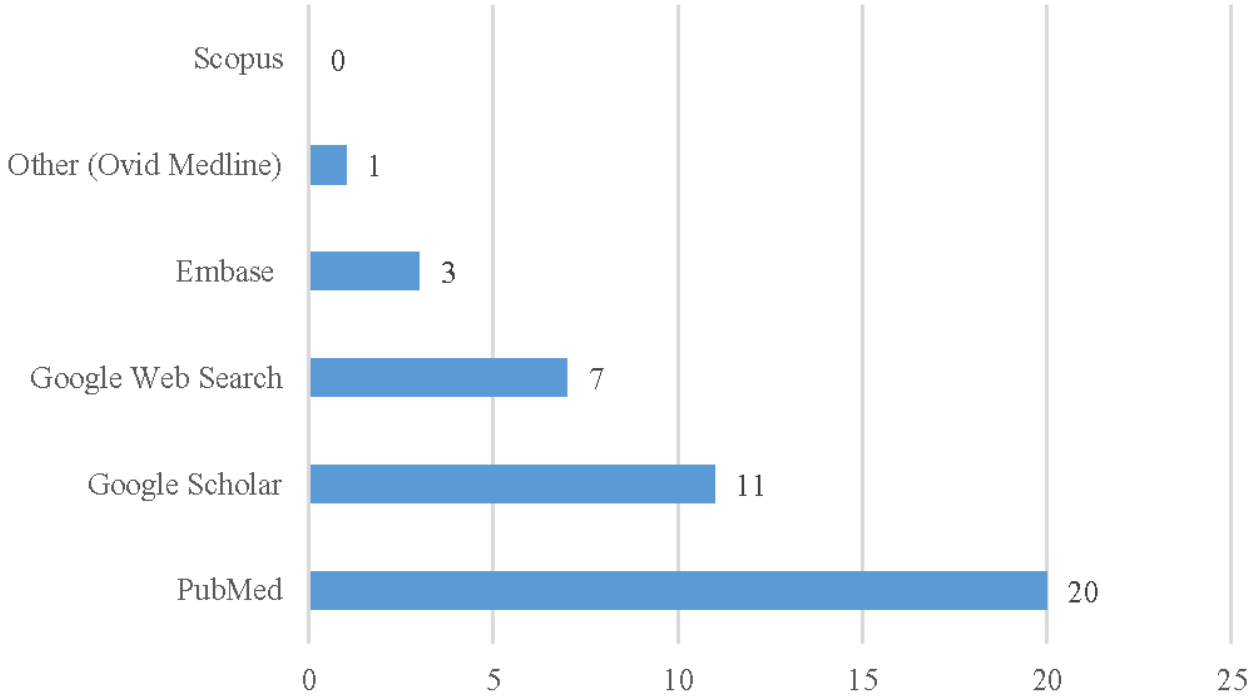
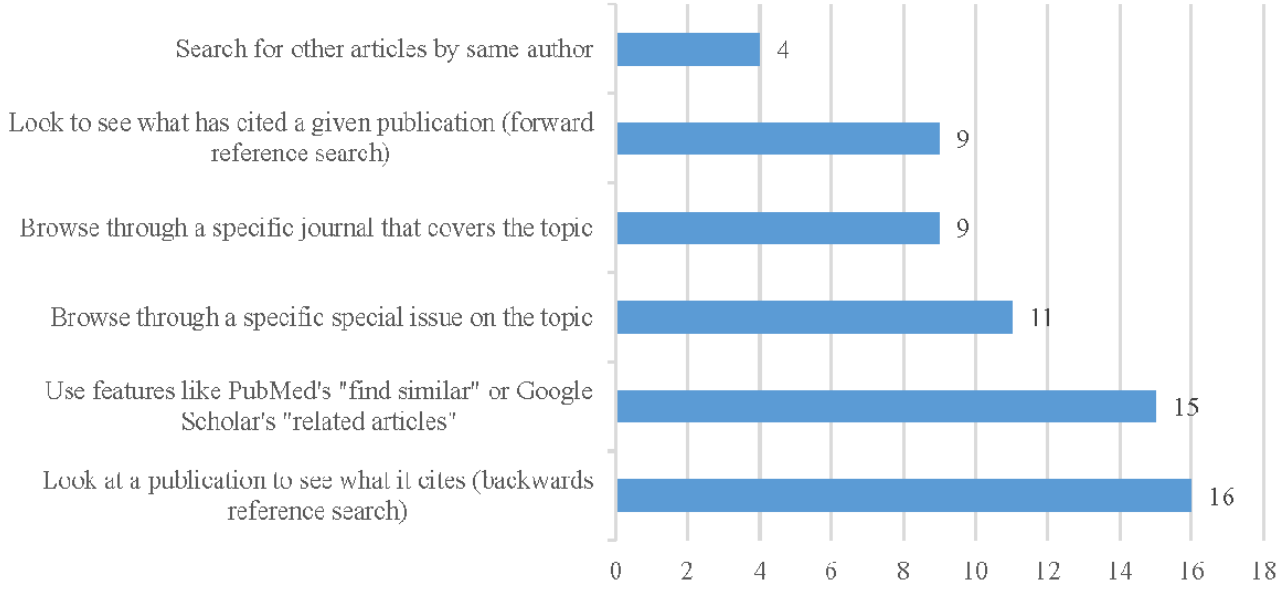


Figure 2. Number of participants reporting using various search strategies N=20.



Additional search strategies used by participants included implementing backward reference searches (16/20, 80%) where additional publications were identified from a paper's bibliography (Figure 2). Features like "find similar" or "related articles" within search engines were frequently used (15/20, 75%). Browsing through special issues (11/20, 55%) or specific journals (9/20, 45%), and examining a citation index to discover more recent papers that have cited a publication (9/20, 45%) were less common strategies. Only 4 respondents (4/20, 20%) reported using author searches as a strategy. Two respondents (2/20, 10%) provided other search strategies used—review table of content alerts for frequently cited journals, and browse safety-related (eg, like MedWatch) notices from countries outside the United States.

Keywords for Searches

Participants identified which types of keywords were most useful when assessing 6 different topics of the PDDI information—(1) that a PDDI exists; (2) seriousness; (3) clinical consequences; (4) management options; (5) mechanism; and (6) health outcomes. Across all 6 PDDI topics, a drug's generic name was the most common keyword used, mentioned 118 times (Table 2). All 20 respondents (20/20, 100%) used the generic name in the first 4 of these PDDI topics, and 95% (19/20) for the last 2 topics. Other common keywords across

all 6 PDDI topics included using the drug class, selected 70 times and used by 50% (10/20)-65% (13/20) of respondents, enzyme name or identifiers, and transporter name or identifiers, both selected 68 times and used by 50% (10/20)-95% (19/20) of respondents. The use of keywords across the 6 PDDI topics was fairly consistent. However, the keyword "drug interaction" was used more commonly to identify that a PDDI exists (by 15/20, 75% of respondents) compared with the use addressing other PDDI topics (30%, 6/20-50%, 10/20 of responders). Types of keywords involving transporter names or identifiers, enzyme names or identifiers, and pharmacological pathway names or identifiers were most often used in assessing the mechanism of a PDDI and least often for assessing health outcomes.

Two PDDI topics had the highest frequency of keywords used—there were 120 mentions of various types of keywords to identify that a PDDI exists, and 113 for assessing the mechanism of a PDDI. The fewest types of keywords (76) were mentioned for use in assessing health outcomes associated with a PDDI. There were 17 responses to "other keywords" that participants reported using for PDDI searches:

1. Pharmacokinetics
2. Pharmacodynamic outcome
3. Clinical trial
4. Case report
5. Human
6. Metabolism
7. Drug level
8. CYP
9. P-gp
10. Custom list of perpetrators [the precipitant drug of an interaction]
11. Custom list of substrates [the object drug of an interaction]
12. Target side effect
13. QT prolongation
14. Arrhythmia
15. Arrhythmic
16. Torsades
17. Herb-drug interactions

Some keywords specified interaction and study types, while others identified specific mechanisms, or targeted particular drugs, diagnoses, or adverse events associated with PDDIs. The keywords below are the actual terms entered in the survey by participants (Table 2). It is worth noting that terms like "precipitants" (also known as perpetrators—meaning the drug that causes the interaction) and "object drugs" (also known as victims—meaning the drug that is affected in the interaction) may not be universally recognized, as there is no standard nomenclature for the role of each drug in the interacting pair.

Study Types for Searches

A similar approach was used to assess the study types respondents include and exclude in their searches to address the same 6 PDDI topics. Trials were the most frequent type of study respondents reported including in searches for PDDI evidence (Table 3). Case reports and systematic reviews were the next most common study designs. Review papers and case series were the least common study designs reported, but were used by 55% (11/20) and 85% (17/20) of respondents, respectively. Overall, a greater variety of study types were used to search for evidence that a PDDI exists compared with the other PDDI topic areas. The fewest study types were used in searches assessing management options for PDDIs. Overall, 6 responses were provided by respondents for "other study types included in PDDI searches" including retrospective, dose-effect relationship, pharmacokinetic, animal data, meeting abstracts, and textbooks.

Table 2. Types of keywords used when searching for different areas of the potential drug-drug interaction information.

Keyword types used	PDDI[a] topics assessed					
	Existence of PPDIs, n (%)	Seriousness of PDDIs, n (%)	Clinical consequences of PDDIs, n (%)	Options to manage PDDIs, n (%)	Mechanism of a PDDI, n (%)	Health outcomes of a PDDI, n (%)
Generic name	20 (100)	20 (100)	20 (100)	20 (100)	19 (95)	19 (95)
Drug class	13 (65)	11 (55)	13 (65)	12 (60)	11 (55)	10 (50)
Enzyme name or identifiers	16 (80)	10 (50)	9 (45)	10 (50)	17 (85)	6 (30)
Transporter names or identifiers	15 (75)	10 (50)	10 (50)	9 (45)	19 (95)	5 (25)
Keyword "drug interaction"	15 (75)	7 (35)	8 (40)	10 (50)	10 (50)	6 (30)
Ingredient names	9 (45)	8 (40)	8 (40)	8 (40)	8 (40)	8 (40)
Pharmacological pathway names or identifiers	10 (50)	6 (30)	7 (35)	5 (25)	13 (65)	4 (20)
Brand name	6 (30)	6 (30)	6 (30)	7 (35)	5 (25)	6 (30)
Drug product name	5 (25)	5 (25)	6 (30)	6 (30)	4 (20)	5 (25)
Drug identifiers	5 (25)	4 (20)	5 (25)	4 (20)	2 (10)	4 (20)
Specific author names	2 (10)	0 (0)	2 (10)	0 (0)	1 (5)	1 (5)

[a]PPDI: potential drug-drug interaction.

Table 3. Study types included in potential drug-drug interaction searches.

Study types included to assess	PDDI[a] topics assessed					
	Existence of PPDIs, n (%)	Seriousness of PDDIs, n (%)	Clinical consequences of PDDIs, n (%)	Options to manage PDDIs, n (%)	Mechanism of a PDDI, n (%)	Health outcomes of a PDDI, n (%)
Trials	18 (90)	18 (90)	17 (85)	14 (70)	16 (80)	16 (80)
Case reports	19 (95)	16 (80)	16 (80)	12 (60)	12 (60)	13 (65)
Systematic reviews	17 (85)	14 (70)	15 (75)	12 (60)	15 (75)	14 (70)
Meta-analyses	15 (75)	15 (75)	15 (75)	11 (55)	12 (60)	14 (70)
Review papers	15 (75)	11 (55)	13 (65)	13 (65)	15 (75)	13 (65)
Case series	17 (85)	14 (70)	15 (75)	11 (55)	12 (60)	11 (55)

[a]PPDI: potential drug-drug interaction.

Table 4. Study types excluded in potential drug-drug interaction searches.

Study types excluded to assess	PDDI[a] topics assessed					
	Existence of PPDIs, n (%)	Seriousness of PDDIs, n (%)	Clinical consequences of PDDIs, n (%)	Options to manage PDDIs, n (%)	Mechanism of a PDDI, n (%)	Health outcomes of a PDDI, n (%)
Animal	10 (50)	15 (75)	17 (85)	16 (80)	8 (40)	15 (75)
In vitro inhibition of enzyme	5 (25)	12 (60)	15 (75)	14 (70)	1 (5)	14 (70)
In vitro inhibition of transporter	5 (25)	12 (60)	15 (75)	14 (70)	1 (5)	14 (70)
In vitro substrate of enzyme	4 (20)	11 (55)	15 (75)	14 (70)	1 (5)	13 (65)
In vitro substrate of transporter	4 (20)	11 (55)	15 (75)	14 (70)	1 (5)	13 (65)
Meeting abstracts	3 (15)	5 (25)	5 (25)	5 (25)	3 (15)	3 (15)
Conference proceedings	3 (15)	3 (15)	3 (15)	3 (15)	4 (20)	1 (5)
Case reports	1 (5)	1 (5)	0 (0)	0 (0)	1 (5)	1 (5)
Case series	0 (0)	1 (5)	0 (0)	0 (0)	1 (5)	0 (0)

[a]PPDI: potential drug-drug interaction.

When asked what types of evidence were excluded from search strategies, respondents reported that animal and *in vitro* studies were not frequently included in the evidence base (Table 4). Very few respondents excluded studies for data relating to the mechanism of action of PDDIs or identifying whether a PDDI exists. Highest numbers of study types excluded in searches were for assessing clinical consequences and management options of PDDIs. Across the 6 PDDI search topics, animal studies were excluded most often, while case reports and case series were rarely excluded.

Data Resources for Searches

Subscription databases most commonly used by participants when searching for PDDI information were Lexicomp (9/20, 45%) and Micromedex DRUG-REAX (8/20, 40%; Figure 3). Seven other subscription databases were mentioned by one respondent each (1/20, 5%)—YouScript/Genelex, Wolters Kluwer, VA CPRS, Medi-Span, Clinical Pharmacology, e-answers, and www.naturaldatabase.com. Open-access databases used most commonly were Drugs@FDA (17/20, 85%), DailyMed (13/20, 65%), and Flockhart Tables (12/20, 60%; Figure 4). None of the respondents reported using the Merck Manual. Other open-access databases reported included CredibleMeds for QT info/AZ CERT QT Meds by respondents (2/20, 10%), www.naturaldatabase.com, drugs.com, product labels, and www.fungalpharmacology.org/tool used by one respondent each (1/20, 5%).

Participants reported using a variety of compendia including Facts and Comparisons (8/20, 40%), Top 100 Drug Interactions (7/20, 35%), Drug Interactions: Analysis and Management (6/20, 30%), and American Hospital Formulary Service Drug Information (6/20, 30%; Figure 5). Other compendia mentioned by one participant each (1/20, 5%) included the VA CPRS and Clinical Pharmacology. The most commonly reported Web-based resources were product labels (18/20, 90%), MedWatch, and DailyMed (both 12/20, 60%; Figure 6). In addition, 50% (10/20) of participants reported using drug manufacturers' websites or contacting them directly for information. Of other Web-based resources, the least utilized was the Agency for Healthcare Research and Quality Effective Healthcare website (1/20, 5%) and the Drug Effectiveness Review Project (DERP; 0/20, 0%). Participants mentioned several Web-based resources in addition to the ones that the survey provided, including Credible Meds, Product non-disclosure agreements, guidelines for managing interactions, summary of product characteristics from European Medicines Agency, Spain Agency of Medicines and Medical Devices, British electronic Medicines Compendium, Australian

Therapeutic Goods Administration, and Canada Drug Product Database, Electronic Medicines Compendium, Therapeutic Goods Administration, and Food and Drug Administration. One resource, naturaldatabase.com, was listed by 1 participant (1/20, 5%) as a subscription resource, while a different participant considered it an open-access resource.

When asked if they would be willing to share one of their search phrases used in PDDI evidence search strategies, 7 participants (7/20, 35%) provided examples. Search phrases typically included the drug names and the term "drug interaction." Several named specific study types or the inhibition pathway involved in the interaction. One example specified the following PubMed search: (drug-drug interaction AND ((Clinical Trial[publication type] OR Case Reports[publication type]) AND Humans[Mesh])). See Multimedia Appendix 2 for all 7 searches.

Six respondents provided additional comments regarding their search strategies for PDDI evidence. Several listed data sources that are the highest priority or specified custom lists they use in searches. One respondent noted:

> *Because PDDI data are often used in medical-legal cases, I will sometimes search for legal precedent (ie, has the strength of the evidence supporting a clinically significant PDDI survived a challenge in court).*

Multimedia Appendix 3 lists all 6 comments.

Figure 3. The number of participants reporting using subscription databases when conducting potential drug-drug interaction (PDDI) searches (N=20). UW: University of Washington.

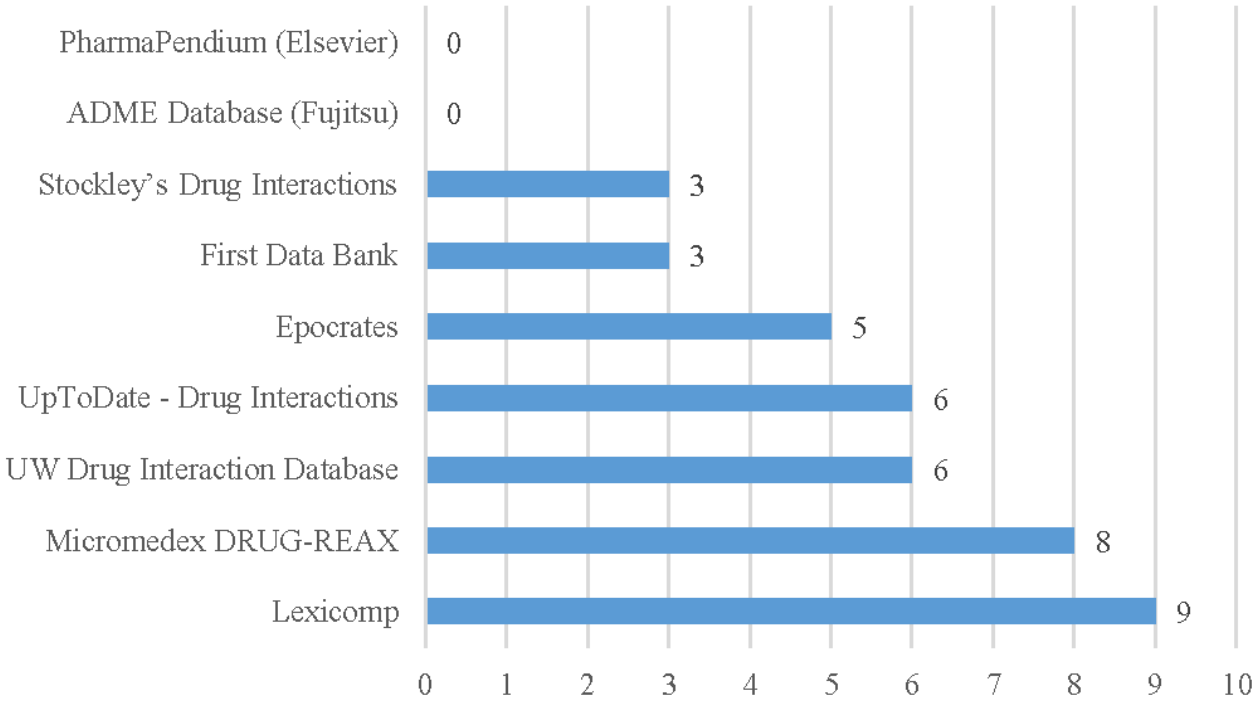


Figure 4. Number of participants reporting using open access databases when conducting PDDI searches N=20.

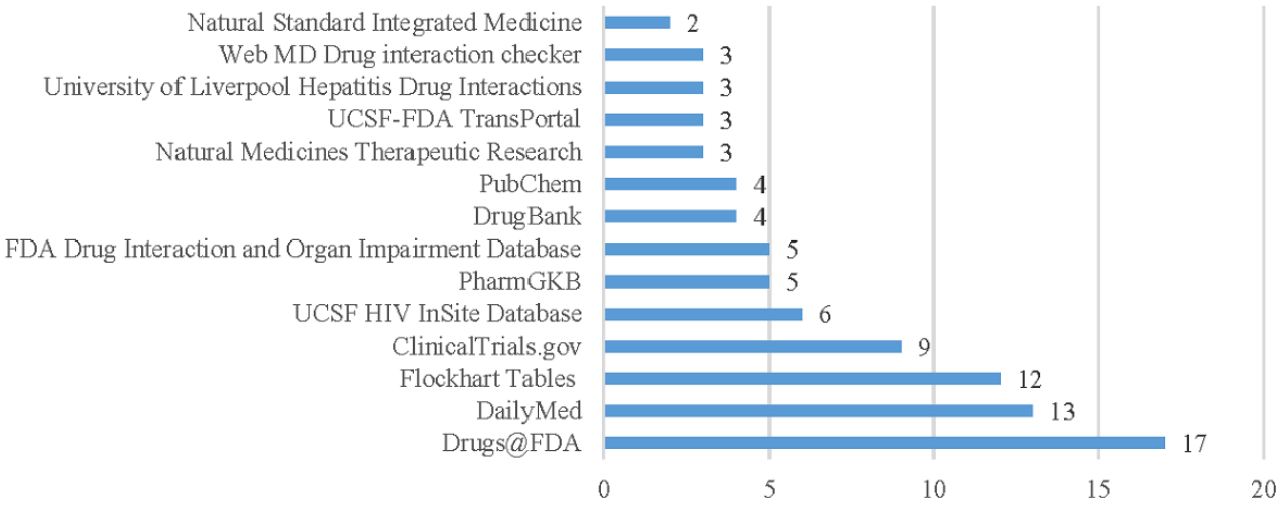


Figure 5. The number of participants reporting using compendia when conducting potential drug-drug interaction (PDDI) searches (N=20).

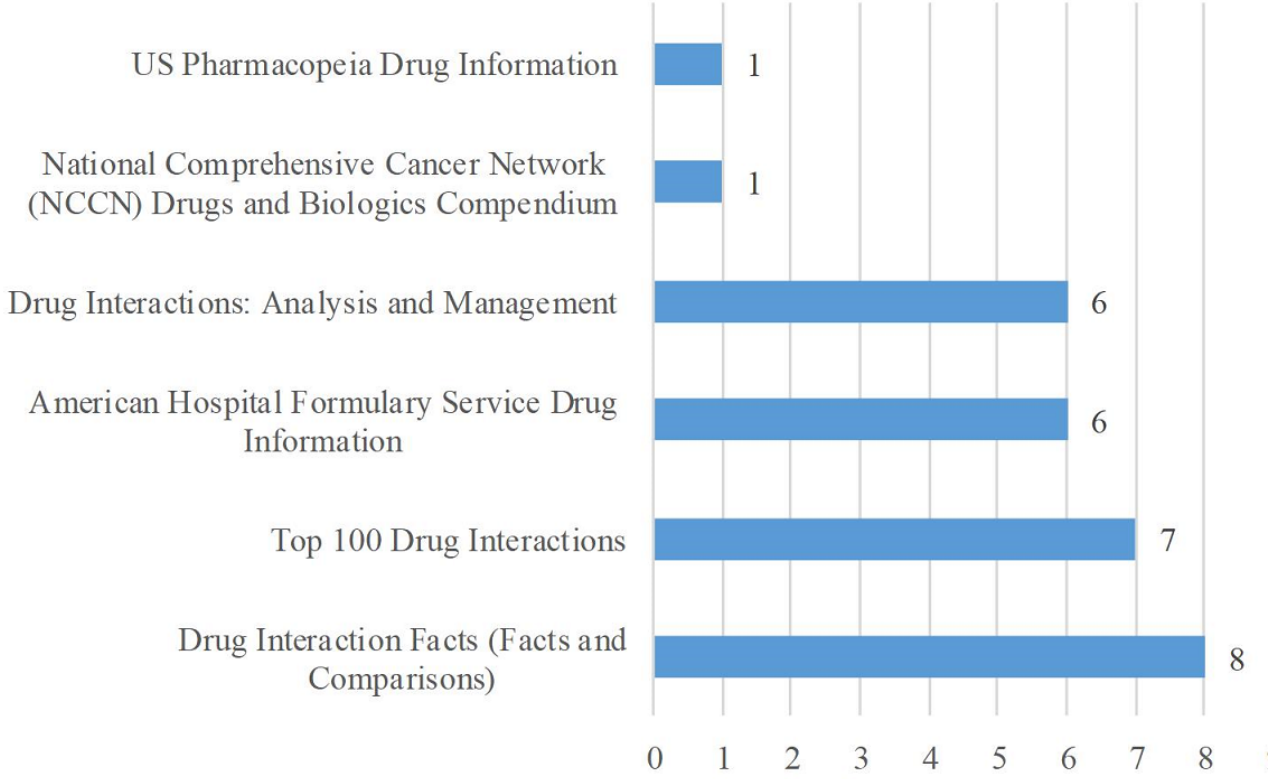
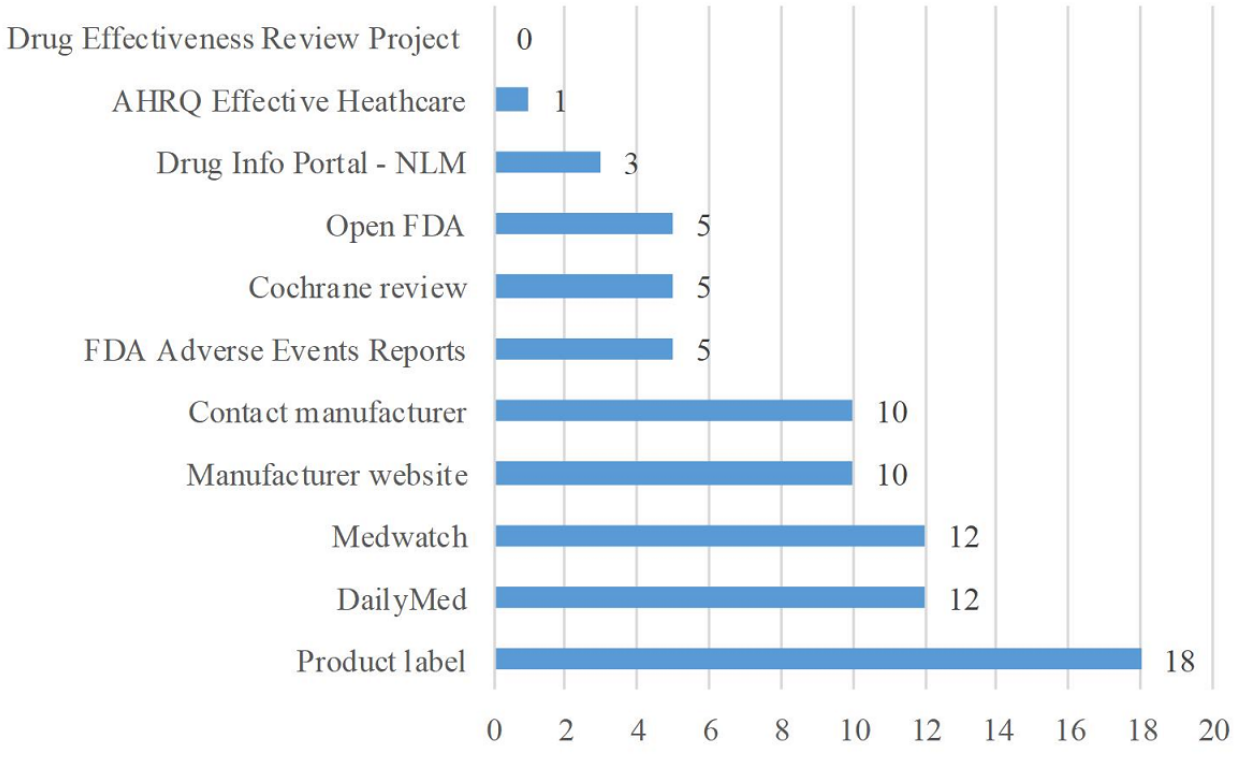


Figure 6. The number of participants reporting using Web-based resources when conducting potential drug-drug interaction (PDDI) searches (N=20). AHRQ: Agency for Healthcare Research and Quality; FDA: Food and Drug Administration; NLM: National Library of Medicine.



Discussion

Principal Findings

This study sought to assess the search strategies used by drug interaction experts when searching for PDDI evidence. Textbox 1 summarizes the key findings and recommendations from this study. We found that among drug interaction experts there are some consistent uses of keywords (eg, generic name used by all respondents), search engines (eg, PubMed used by all), databases (eg, Drugs@FDA used by 17/20, 85%), and Web resources (eg, product labels used by 19/20, 90%). However, a variety of resources are used to search for the PDDI information with little consistency across experts. For example, 14 subscription databases and 19 open-access databases were used by 5% (1/20)-45% (9/20) of participants. In addition, 8 compendia were used by 5% (1/20)-25% (5/20) of respondents. Similarly, study types were not standard across the various PDDI topic searches.

Findings from our survey indicate that experts develop and conduct searches without assistance from a librarian, even though a librarian can play a valuable role in setting up search filters and determining the most accurate terms to use. This might be attributed to the lack of access to library services. It could also be from the experts' confidence that they have the competency to perform their own queries, particularly in their specific area of expertise. All participants use PubMed and over half use Google tools when searching for papers. Google Scholar differs from PubMed, using a different algorithm to produce a broader domain, including gray literature such as conference proceedings, doctoral theses, white papers, etc. Using PubMed and Google Scholar together might result in a more comprehensive search. One expert (1/20, 5%) reported using Ovid MEDLINE, a proprietary database using the same underlying data as PubMed. The abstracting service, Scopus, was not used at all, and few experts reported using EMBASE, which has a more international focus than PubMed. This could be attributed to a lack of familiarity or simply a resource issue as EMBASE is proprietary.

This study found that drug interaction experts use a variety of keyword strategies and evidence sources while searching for the PDDI information. Not surprisingly, transporter names or identifiers, enzyme names or identifiers, and pharmacological pathway names or identifiers were the most often used types of keywords for assessing the mechanism of a PDDI and least often for assessing health outcomes. The topic of PDDI health outcomes stood out as having the least number of terms that experts considered relevant with 14 fewer keyword types selected than the second lowest-ranking topic, PDDI seriousness. Possible explanations might be that health outcomes are highly variable and require highly specific keywords that we did not include as selections, or this PDDI topic may be irrelevant to most of these drug interaction experts. Only a small number of participants contributed "other" keywords such as "QT prolongation," "Torsades," and "Arrhythmia" that are relevant health outcomes.

A list of 17 "other keywords" provided by these experts indicates tremendous variation in searching the literature. As expected, participants selected a variety of keyword types as relevant for searches depending on the specific PDDI information being sought. It is reasonable that different terms would be appropriate for different searches, that is, we would expect different keywords to be used to assess the mechanism compared with clinical consequences associated with a PDDI. The research question should ultimately drive the search. This makes the requirements of a standardized PDDI search strategy more complex.

Textbox 1. Key study findings and recommendations.

Consistent search strategies among drug interaction experts

- Keywords: Generic name used to search 6 potential drug-drug interaction (PDDI) topics (19/20, 95%)
- Search engines: PubMed used (20/20, 100%)
- Databases: Drugs@FDA used (17/20, 85%)
- Web resources: Product labels used (18/20, 90%)
- Librarian assistance: Not used 90% (18/20)-95% (19/20)

Inconsistent search strategies among drug interaction experts

- Keywords: Variation across 6 PDDI topics
- Databases: Use of 14 subscription and 19 open access
- Compendia: Use of 8 compendia
- Study types: Variation across the 6 PDDI topics

Recommendations to improve PDDI evidence search strategies

- Develop validated search strategies
- Integrate PDDI-relevant automation in search tools more effectively
- Use multiple search engines
- Involve librarian assistance

Although participants were not specifically asked whether they use Boolean operators to fine tune search strings, nearly all example search phrases provided by participants included AND/OR Boolean terms, as well as modifiers such as quotation marks, to indicate exact phrases, and brackets around OR statements. These Boolean strategies (as well as an asterisk following a root word to capture other variations of the term) can be utilized to help create specific search strings that save time in filtering results. Ensuring that experts know how to use these tools appropriately can help target search results for identifying PDDI evidence.

With respect to study types, participants most often indicated that trials, case reports, and systematic reviews were relevant to their PDDI searches, with some variation depending on the PDDI topic. Interestingly, the relative importance of the study types among those surveyed does not reflect their relative abundance in PubMed. For example, participants indicated using the review study types less than the trial study type across every PDDI topic. However, at the time of this writing, a PubMed PDDI review search returns more papers than either a clinical trial search or a case report search:

- review search (16,022 results): ("drug interactions"[MeSH Terms]) AND ("review"[Publication Type]
- clinical trial search (10,789 results): ("drug interactions"[MeSH Terms]) AND ("clinical trial"[Publication Type])
- case report search (5962 results): ("drug interactions"[MeSH Terms]) AND ("case reports"[Publication Type])

Animal and *in vitro* data study types were generally considered nonrelevant. Between 75% (15/20) and 85% (17/20) of respondents do not exclude meeting abstracts and conference proceedings, suggesting that most participants are willing to consider a wide variety of PDDI evidence sources. These less rigorous sources may be used to assess early warnings of possible interactions but may be excluded when searching for well-established data to support decision making. Participants who excluded abstracts and conference proceedings might do so because these evidence types generally do not result in a published study or might not include key information about PDDIs.

The subscription database use was reported by less than half (5% [1/20]-45% [9/20]) of the respondents. Although respondents were drug information experts, many might not have access to or be aware of these databases. In contrast, >50% of participants indicated using 3 open-access sources, Drugs@FDA (17/20, 85%), DailyMed (13/20, 65%), and Indiana University P450 Drug Interaction Tables (Flockhart Tables; 12/20, 60%). Not surprisingly, the Merck Manual was not used by any of the experts. In our experience, this publication has fallen out of favor over the past several decades as more contemporary Web-based resources have become available. No single drug information compendium had broad usage by participants. In terms of Web resources, the survey identified reliance on manufacturer information—90% (18/20) use product labels, 50% (10/20) use company websites, and 50% (10/20) contact companies for information. Of interest, 13 Web-based resources were used by <5 participants. This finding might indicate that several potentially relevant information sources are not broadly known to drug interaction experts. Alternatively, the research questions addressed by these experts may best be answered by this subset of 13 Web-based resources.

Comparison of the Results With Prior Work

A recent survey of health care information professionals identified the need for assistance in developing complex search strategies, especially if searches and results are to be transparent and repeatable [31]. In addition, respondents wanted to increase the specificity of searches to minimize the number of nonrelevant papers. These findings support our recommendation that the standardization and enhanced functionality are areas that need further improvement for search strategy development.

Techniques to improve literature searches are evolving. For example, Duda et al assessed PDDI queries of PubMed using the standard Boolean query method and a novel machine learning method [28]. They developed a reference set of 2000 titles and abstracts (published between 1985 and 2002) discussing PDDI studies; 10% (200/2000) involving interactions between the drug pairs and 90% (1800/2000) containing general information about PDDIs. They then identified the sensitivity, specificity, and positive predictive value of the 2 PubMed queries. The performance of a novel machine classifier (trained on titles, abstracts, and MeSH headings) was found to be comparable to that of the queries.

Other studies have focused on data mining to identify PDDIs within titles, abstracts, and papers [16-28]. Included in these studies are approaches to use various kinds of machine learning, including linear kernels (eg, Support Vector Machines) [18,19,32], nonlinear kernels (eg, Graph Models) [22], random forest [16], various neural network architectures [17,21,26,33], advanced use of linguistic parts of speech and linguistic features [19,23], unsupervised topical models [25], and semantic features from terminologies or ontologies [16,27,32,34,35]. In general, the goal of these sophisticated approaches is to accurately extract PDDI data from a large body of scientific literature. Two different formal computing challenges have focused on the same topic [36,37]. As a whole, these studies show that greater automation during the literature search task is feasible. In principle, it would be possible for an individual or organization to implement any of the published algorithms within a custom search portal. However, none of the participants indicated they used specialized tools in their search activities. To the best of our knowledge, the major scientific literature search engines (eg, PubMed, Google Scholar) do not currently implement any of these advanced methods. Future work focusing on disseminating these advanced searching techniques is a logical step to improve search strategies for identifying PDDI evidence.

Although individual professions (eg, academic, legal, medical, governmental, and pharmacy) differ in their searching needs, there is a common focus on identifying information sources, accessing appropriate systems, and managing knowledge. The First International Workshop on Professional Search was held in 2018 [38], which highlighted that requirements for professional search tasks differ from those of generic Web search engines. In professional searching, it is important to

identify information needs, behavioral patterns, and understand the interface between the user and the information system being utilized. Hence, more research is needed in the field of professional search to assess synergies across professions.

Limitations

This study has several limitations. Participants were sampled by convenience from individuals who we previously identified as conducting PDDI evidence search and synthesis for work. Participants were predominantly English-speaking and from the United States. We are unable to speculate about the generalizability of the results to other locales. This study focuses only on one population, drug interaction experts, and cannot be generalized to how other populations, such as clinicians, conduct searches for PDDI information. However, it may be reasonable to assume that the lack of consistency in search approaches among experts would extend to other, less experienced professionals searching for PDDI evidence. It could be argued that the survey response rate was relatively low (20/70, 29%). However, drug interaction experts whose work involved PDDIs are a specialized population, and the sampling frame was cast relatively wide (n=70).

Inherent to surveys is whether respondents interpreted the question as intended. For example, questions about searching for the topic of whether a PDDI exists could have been interpreted broadly by participants to include the topics of mechanism, seriousness, and clinical consequences rather than the single topic of existence intended. In addition, some participants might have perceived differences between identifying that a PDDI exists and determining whether it is clinically meaningful. A similar limitation could apply for study types. For example, participants indicated case series as less relevant in searches than case reports across all 6 PDDI topics. This is counterintuitive if case series are considered to be the reporting of multiple case studies, which should increase the credibility and usefulness of case series. Based on participants' responses, case series may not be well defined.

We chose to use different lists of study types in the 2 questions that asked participants about included and excluded study types. While this precludes direct comparisons of all responses regarding included and excluded study types, it was done for several reasons. First, 5 study types were used in both lists, and these show reciprocal results. By selecting different study types, we could be more specific when listing study types that would be more likely to be selected as included or excluded types. Finally, the partial duplication of study types limited the survey length.

There is a lack of clearly specified search strategies that have been validated for retrieving the most complete and precise PDDI evidence possible. These standard search strategies would simplify the search process, saving drug interaction experts time and energy. Moreover, it would ensure that important sources for PDDI evidence are not overlooked. The information would be more comprehensive, with the goal of limiting discrepancies across data sources, thereby reducing confusion and frustration among end users. In future work, we plan to develop candidate search strategies to assess whether the recommended standards for PDDI evidence searches are useful.

Conclusions

In conclusion, drug interaction experts appear to use varying keyword strategies, databases, and Web resources when seeking to identify PDDI evidence. This study supports the need to create comprehensive search strategies to identify relevant PDDI evidence. Incorporating automated tools may enhance the ability to locate, synthesize, and apply the PDDI information. Future research is needed to improve the existing search tools, develop standards for search strategy recommendations, and evaluate their usefulness and accuracy in identifying the relevant PDDI evidence.

Acknowledgments

This work was supported by grant R01LM011838 from the US National Library of Medicine, National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey instrument.

[PDF File (Adobe PDF File), 80KB - jmir_v21i1e11182_app1.pdf]

Multimedia Appendix 2

Example search strings.

[PDF File (Adobe PDF File), 23KB - jmir_v21i1e11182_app2.pdf]

Multimedia Appendix 3

Respondent comments.

[PDF File (Adobe PDF File), 14KB - jmir_v21i1e11182_app3.pdf]

References

1. Dechanont S, Maphanta S, Butthum B, Kongkaew C. Hospital admissions/visits associated with drug-drug interactions: a systematic review and meta-analysis. Pharmacoepidemiol Drug Saf 2014 May;23(5):489-497. [doi: 10.1002/pds.3592] [Medline: 24616171]
2. Institute of Medicine. In: Aspden P, Wolcott J, Bootman JL, Cronenwett LR, editors. Preventing Medication Errors. Washington DC: The National Academies Press; 2007.
3. Böttiger Y, Laine K, Andersson ML, Korhonen T, Molin B, Ovesjö M, et al. SFINX-a drug-drug interaction database designed for clinical decision support systems. Eur J Clin Pharmacol 2009 Jun;65(6):627-633. [doi: 10.1007/s00228-008-0612-5] [Medline: 19205683]
4. Bryant AD, Fletcher GS, Payne TH. Drug interaction alert override rates in the Meaningful Use era: no evidence of progress. Appl Clin Inform 2014;5(3):802-813 [FREE Full text] [doi: 10.4338/ACI-2013-12-RA-0103] [Medline: 25298818]
5. Abarca J, Malone DC, Armstrong EP, Grizzle AJ, Hansten PD, Van BRC, et al. Concordance of severity ratings provided in four drug interaction compendia. J Am Pharm Assoc (2003) 2004;44(2):136-141. [Medline: 15098847]
6. van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc 2006;13(2):138-147 [FREE Full text] [doi: 10.1197/jamia.M1809] [Medline: 16357358]
7. McEvoy DS, Sittig DF, Hickman T, Aaron S, Ai A, Amato M, et al. Variation in high-priority drug-drug interaction alerts across institutions and electronic health records. J Am Med Inform Assoc 2017 Dec 01;24(2):331-338 [FREE Full text] [doi: 10.1093/jamia/ocw114] [Medline: 27570216]
8. Saverno KR, Hines LE, Warholak TL, Grizzle AJ, Babits L, Clark C, et al. Ability of pharmacy clinical decision-support software to alert users about clinically important drug-drug interactions. J Am Med Inform Assoc 2011;18(1):32-37 [FREE Full text] [doi: 10.1136/jamia.2010.007609] [Medline: 21131607]
9. Weingart SN, Simchowitz B, Shiman L, Brouillard D, Cyrulik A, Davis RB, et al. Clinicians' assessments of electronic medication safety alerts in ambulatory care. Arch Intern Med 2009 Sep 28;169(17):1627-1632. [doi: 10.1001/archinternmed.2009.300] [Medline: 19786683]
10. Fung KW, Kapusnik-Uner J, Cunningham J, Higby-Baker S, Bodenreider O. Comparison of three commercial knowledge bases for detection of drug-drug interactions in clinical decision support. J Am Med Inform Assoc 2017 Jul 01;24(4):806-812. [doi: 10.1093/jamia/ocx010] [Medline: 28339701]
11. Romagnoli KM, Nelson SD, Hines L, Empey P, Boyce RD, Hochheiser H. Information needs for making clinical recommendations about potential drug-drug interactions: a synthesis of literature review and interviews. BMC Med Inform Decis Mak 2017 Dec 22;17(1):21 [FREE Full text] [doi: 10.1186/s12911-017-0419-3] [Medline: 28228132]
12. 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]
13. Beynon R, Leeflang MMG, McDonald S, Eisinga A, Mitchell RL, Whiting P, et al. Search strategies to identify diagnostic accuracy studies in MEDLINE and EMBASE. Cochrane Database Syst Rev 2013 Sep 11(9):MR000022. [doi: 10.1002/14651858.MR000022.pub3] [Medline: 24022476]
14. Leenaars M, Hooijmans CR, van Veggel N, ter Riet G, Leeflang M, Hooft L, et al. A step-by-step guide to systematically identify all relevant animal studies. Lab Anim 2012 Jan;46(1):24-31 [FREE Full text] [doi: 10.1258/la.2011.011087] [Medline: 22037056]
15. Golder S, Loke Y. Search strategies to identify information on adverse effects: a systematic review. J Med Libr Assoc 2009 Apr;97(2):84-92 [FREE Full text] [doi: 10.3163/1536-5050.97.2.004] [Medline: 19404498]
16. Percha B, Garten Y, Altman RB. Discovery and explanation of drug-drug interactions via text mining. Pac Symp Biocomput 2012:410-421 [FREE Full text] [Medline: 22174296]
17. Lim S, Lee K, Kang J. Drug drug interaction extraction from the literature using a recursive neural network. PLoS One 2018;13(1):e0190926 [FREE Full text] [doi: 10.1371/journal.pone.0190926] [Medline: 29373599]
18. Kolchinsky A, Lourenço A, Li L, Rocha LM. Evaluation of linear classifiers on articles containing pharmacokinetic evidence of drug-drug interactions. Pac Symp Biocomput 2013:409-420 [FREE Full text] [Medline: 23424145]
19. Bui Q, Sloot PM, van Mulligen EM, Kors JA. A novel feature-based approach to extract drug-drug interactions from biomedical text. Bioinformatics 2014 Dec 1;30(23):3365-3371 [FREE Full text] [doi: 10.1093/bioinformatics/btu557] [Medline: 25143286]
20. Kolchinsky A, Lourenço A, Wu HY, Li L, Rocha LM. Extraction of pharmacokinetic evidence of drug-drug interactions from the literature. PLoS One 2015;10(5):e0122199 [FREE Full text] [doi: 10.1371/journal.pone.0122199] [Medline: 25961290]
21. Liu S, Tang B, Chen Q, Wang X. Drug-Drug Interaction Extraction via Convolutional Neural Networks. Comput Math Methods Med 2016;2016:6918381 [FREE Full text] [doi: 10.1155/2016/6918381] [Medline: 26941831]
22. Zheng W, Lin H, Zhao Z, Xu B, Zhang Y, Yang Z, et al. A graph kernel based on context vectors for extracting drug-drug interactions. J Biomed Inform 2016 Dec;61:34-43 [FREE Full text] [doi: 10.1016/j.jbi.2016.03.014] [Medline: 27012903]

23. Bokharaeian B, Diaz A, Chitsaz H. Enhancing Extraction of Drug-Drug Interaction from Literature Using Neutral Candidates, Negation, and Clause Dependency. PLoS One 2016;11(10):e0163480 [FREE Full text] [doi: 10.1371/journal.pone.0163480] [Medline: 27695078]
24. Barneh F, Jafari M, Mirzaie M. Updates on drug-target network; facilitating polypharmacology and data integration by growth of DrugBank database. Brief Bioinform 2016 Dec;17(6):1070-1080. [doi: 10.1093/bib/bbv094] [Medline: 26490381]
25. Li D, Liu S, Rastegar-Mojarad M, Wang Y, Chaudhary V, Therneau T, et al. A Topic-modeling Based Framework for Drug-drug Interaction Classification from Biomedical Text. AMIA Annu Symp Proc 2016;2016:789-798 [FREE Full text] [Medline: 28269875]
26. Zheng W, Lin H, Luo L, Zhao Z, Li Z, Zhang Y, et al. An attention-based effective neural model for drug-drug interactions extraction. BMC Bioinformatics 2017 Oct 10;18(1):445 [FREE Full text] [doi: 10.1186/s12859-017-1855-x] [Medline: 29017459]
27. Fiszman M, Rindflesch TC, Kilicoglu H. Summarizing drug information in Medline citations. AMIA Annu Symp Proc 2006:254-258 [FREE Full text] [Medline: 17238342]
28. Duda S, Aliferis C, Miller R, Statnikov A, Johnson K. Extracting drug-drug interaction articles from MEDLINE to improve the content of drug databases. AMIA Annu Symp Proc 2005:216-220 [FREE Full text] [Medline: 16779033]
29. Payne TH, Hines LE, Chan RC, Hartman S, Kapusnik-Uner J, Russ AL, et al. Recommendations to improve the usability of drug-drug interaction clinical decision support alerts. J Am Med Inform Assoc 2015 Nov;22(6):1243-1250. [doi: 10.1093/jamia/ocv011] [Medline: 25829460]
30. Scheife RT, Hines LE, Boyce RD, Chung SP, Momper JD, Sommer CD, et al. Consensus recommendations for systematic evaluation of drug-drug interaction evidence for clinical decision support. Drug Saf 2015 Feb;38(2):197-206 [FREE Full text] [doi: 10.1007/s40264-014-0262-8] [Medline: 25556085]
31. Russell-Rose T, Chamberlain J. Expert Search Strategies: The Information Retrieval Practices of Healthcare Information Professionals. JMIR Med Inform 2017 Oct 02;5(4):e33 [FREE Full text] [doi: 10.2196/medinform.7680] [Medline: 28970190]
32. Boyce R, Gardner G, Harkema H. Using natural language processing to extract drug-drug interaction information from package inserts. In: BioNLP: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing. Montréal, Canada: Association for Computational Linguistics; 2012 Presented at: Workshop on Biomedical Natural Language Processing; June 8, 2012; Montréal, Canada p. 206-213 URL: http://www.aclweb.org/anthology/W12-2426
33. Suárez-Paniagua V, Segura-Bedmar I, Martínez P. Exploring convolutional neural networks for drug-drug interaction extraction. Database (Oxford) 2017 Dec 01;2017:1-15 [FREE Full text] [doi: 10.1093/database/bax019] [Medline: 28605776]
34. Noor A, Assiri A, Ayvaz S, Clark C, Dumontier M. Drug-drug interaction discovery and demystification using Semantic Web technologies. J Am Med Inform Assoc 2017 May 01;24(3):556-564. [doi: 10.1093/jamia/ocw128] [Medline: 28031284]
35. Lu Y, Figler B, Huang H, Tu Y, Wang J, Cheng F. Characterization of the mechanism of drug-drug interactions from PubMed using MeSH terms. PLoS One 2017;12(4):e0173548 [FREE Full text] [doi: 10.1371/journal.pone.0173548] [Medline: 28422961]
36. Segura-Bedmar I, Martínez P, Sánchez-Cisneros D. The 1st DDIExtraction-2011 challenge task:Extraction of drug-drug interactions from biomedical texts. In: The 1st Challenge Task on Drug-Drug Interaction Extraction 2011. 2011 Presented at: 1st Challenge Task on Drug-Drug Interaction Extraction 2011; September 7, 2011; Huelva, Spain p. 1-9.
37. Segura-Bedmar I, Martínez P, Herrero-Zazo M. Lessons learnt from the DDIExtraction-2013 Shared Task. J Biomed Inform 2014 Oct;51:152-164 [FREE Full text] [doi: 10.1016/j.jbi.2014.05.007] [Medline: 24858490]
38. Dietz L, Koesten L, Verberne S. Proceedings of the First International Workshop on Professional Search (ProfS2018). 2018 Presented at: First International Workshop on Professional Search; July 12, 2018; Ann Arbor, Michigan p. 1-42 URL: http://ceur-ws.org/Vol-2127/

Abbreviations

PDDI: Potential drug-drug interaction

Edited by G Eysenbach; submitted 05.06.18; peer-reviewed by J Chamberlain, T Russell-Rose, P Chai; comments to author 12.07.18; revised version received 05.09.18; accepted 27.09.18; published 04.01.19.

Please cite as:
Grizzle AJ, Horn J, Collins C, Schneider J, Malone DC, Stottlemyer B, Boyce RD
Identifying Common Methods Used by Drug Interaction Experts for Finding Evidence About Potential Drug-Drug Interactions: Web-Based Survey
J Med Internet Res 2019;21(1):e11182
URL: https://www.jmir.org/2019/1/e11182/
doi:10.2196/11182
PMID:30609981

Original Paper

Measuring the Impact of an Open Web-Based Prescribing Data Analysis Service on Clinical Practice: Cohort Study on NHS England Data

Alex J Walker[1], PhD; Helen J Curtis[1], DPhil; Richard Croker[1], MSc; Seb Bacon[1], BA; Ben Goldacre[1], BA, MA, MSc, MBBS, MRCPsych

Evidence Based Medicine DataLab, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Ben Goldacre, BA, MA, MSc, MBBS, MRCPsych
Evidence Based Medicine DataLab
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford, OX2 6GG
United Kingdom
Phone: 44 1865289313
Email: ben.goldacre@phc.ox.ac.uk

Abstract

Background: OpenPrescribing is a freely accessible service that enables any user to view and analyze the National Health Service (NHS) primary care prescribing data at the level of individual practices. This tool is intended to improve the quality, safety, and cost-effectiveness of prescribing.

Objective: We aimed to measure the impact of OpenPrescribing being viewed on subsequent prescribing.

Methods: Having preregistered our protocol and code, we measured three different metrics of prescribing quality (mean percentile across 34 existing OpenPrescribing quality measures, available "price-per-unit" savings, and total "low-priority prescribing" spend) to see whether they changed after the viewing of Clinical Commissioning Group (CCG) and practice pages. We also measured whether practices whose data were viewed on OpenPrescribing differed in prescribing, prior to viewing, compared with those who were not. We used fixed-effects and between-effects linear panel regression to isolate change over time and differences between practices, respectively. We adjusted for the month of prescribing in the fixed-effects model to remove underlying trends in outcome measures.

Results: We found a reduction in available price-per-unit savings for both practices and CCGs after their pages were viewed. The saving was greater at practice level (−£40.42 per thousand patients per month; 95% CI −54.04 to −26.81) than at CCG level (−£14.70 per thousand patients per month; 95% CI −25.56 to −3.84). We estimate a total saving since launch of £243 thosand at practice level and £1.47 million at CCG level between the feature launch and end of follow-up (August to November 2017) among practices viewed. If the observed savings from practices viewed were extrapolated to all practices, this would generate £26.8 million in annual savings for the NHS, approximately 20% of the total possible savings from this method. The other two measures were not different after CCGs or practices were viewed. Practices that were viewed had worse prescribing quality scores overall prior to viewing.

Conclusions: We found a positive impact from the use of OpenPrescribing, specifically for the class of savings opportunities that can only be identified by using this tool. Furthermore, we show that it is possible to conduct a robust analysis of the impact of such a Web-based service on clinical practice.

(J Med Internet Res 2019;21(1):e10929) doi:10.2196/10929

KEYWORDS

drug prescribing; cost control; patient safety; treatment efficacy

Introduction

The OpenPrescribing.net project aims to make prescribing data more accessible and impactful for clinicians, policy makers, and others. It does this through providing a user-friendly Web interface. It is hoped that this will enable safer, more effective, and cost-efficient prescribing by making users more aware of their prescribing behavior through comparisons with peers and by highlighting meaningful changes over time. OpenPrescribing gives a range of specific analyses and tools for each Clinical Commissioning Group (CCG) and individual primary care practice in three broad classes (set out in Textbox 1). Use of OpenPrescribing is driven largely by it being openly accessible and free to use; the service had 70,000 unique users over the past year. Broadly, users are obtained through word of mouth and social media engagement combined with some press coverage. Use of the service is not mandatory or associated with any payment in any locations that we are aware of.

There are many commercial tools that aim to improve prescribing in the United Kingdom. These include Oracle [3], Optum (ScriptSwitch) [4], Prescribing Services [5], and Prescribing Support Services [6]. While these tools may be effective at generating change, there is little publicly available evidence of robust testing. Such testing is an important element of good commissioning practice, in order to ensure that resources are used cost-effectively [7].

We set out to deliver a robust quantitative evaluation of the impact of the use of an open Web-based tool that may act as a template for other evaluations of similar tools, both in terms of the open and reproducible approach and the methodology used. We set out to measure whether any change in prescribing behavior occurs after the use of OpenPrescribing using the most robust observational methods to determine whether there is any causal association. Because it is possible that practices and CCGs who engage with their prescribing data are systematically different to those who do not, we also set out to measure whether practices and CCGs viewing OpenPrescribing already differ from their peers in prescribing behavior prior to using the tool.

Textbox 1. Prescribing Data Available on OpenPrescribing.

1. A range of *prespecified prescribing measures* (eg, "proportion of statins that are high cost" and "antibiotic prescriptions per standardized population unit").
2. A *price-per-unit* tool that uses a novel method developed at OpenPrescribing to identify a range of cost-saving opportunities bespoke for each institution, driven by analysis of national price variation within the same chemical and dose across each month. In essence, we calculate the mean price per unit (comparable to price-per-tablet) for every dose of every drug prescribed, for every practice and Clinical Commissioning Group (CCG); calculate the range of prices nationally; calculate the price per unit at the 10th centile practice or CCG; calculate the available savings for every organization, for every dose of every drug, were they to attain the price per unit of the 10th centile organization; rank these by total savings; and, then, present the highest value prescribing changes for each practice and CCG through a Web interface. Savings are, therefore, identified by a novel method using computationally intensive data analysis across the entire national prescribing dataset. At present, crucially for the analysis presented here, these savings can only be identified using OpenPrescribing [1]. Savings are best realized by viewing practice-level data.
3. A range of tools that collectively calculate the total money spent on items identified by NHS England as a *low priority* on the grounds of extremely low cost-effectiveness [2].

Methods

Prespecification and Protocol Registration

As this is an observational study of the impact of our own service, we endeavored to minimize any potential for conflict of interest impacting on results by fully prespecifying our methods and posting the protocol [8] and analytic code [9] on the Open Science Framework prior to commencement. In the protocol, we specified the outcomes to be measured along with the full analytic approach. The entire analytic code was written against a small sample of 7 institutions' data, and we published it prior to conducting the analysis. There were no substantial changes to the outcomes or methodology (including the analytic code) between prespecification and reporting the results here.

Data and Sources

We obtained the prescribing outcome data from the monthly prescribing dataset published by National Health Service (NHS) digital and aggregated by the OpenPrescribing project. The monthly prescribing datasets contain one row for each treatment and dose for each prescribing organization in NHS primary care in England, describing the number of prescriptions issued and the total cost. Each practice in England belongs to one of 207 CCGs; we aggregated practice data to these CCGs for CCG-level analyses. Practices with a very small list size (<1000) were excluded due to the likelihood of being an atypical practice.

We obtained data on the exposure (CCG and practice page views) from Google Analytics for OpenPrescribing.net. We collected page view data from the launch of the project (December 1, 2015) to the most recent data available at the time of extraction (January 14, 2018). Page view data contain the date of each view, along with information on which practice or CCG's data were viewed and on whether any specific site features were viewed (eg, the price per unit or low-priority features). We aggregated page view dates to month-year.

Exposures

The exposures used for this study are page views on the OpenPrescribing.net site. This is a proxy exposure, as we are not able to attribute site use to an individual practice or CCG, but instead assume that a high proportion of traffic to each practice's or CCG's prescribing page is from persons associated with that practice or CCG.

We generated two different exposure variables to determine the associations with viewing OpenPrescribing. First, for each month, we categorized each practice and CCG into one of the three categories: "not viewed", during "month first viewed", or "after first view", according to when each practice or CCG was viewed on OpenPrescribing in relation that month. We used "not viewed" to classify the months before viewing a CCG or practice. We defined the "month first viewed" as the first month that a practice or CCG page was viewed more than once (in order to exclude one-off visits, which are unlikely to represent full engagement with the site) and "after first view" as every month after the "month first viewed," until the end of follow-up. Second, we generated a variable to describe the total number of views of each practice and CCG page on OpenPrescribing (divided into categories: 0 views and tertiles of the number of practice views among those that were viewed). For the price-per-unit and low-priority outcomes (described below), the exposure was restricted to views of the price-per-unit and low-priority pages for an institution.

Outcomes

We used 3 outcome variables in order to measure the effectiveness of the three main features of OpenPrescribing (see Textbox 1). First, we calculated mean percentile for each of the standard prespecified OpenPrescribing measures [10], excluding the NHS England low-priority measures (which are analyzed separately below) and those where a value judgment is not made (currently direct-acting oral anticoagulants [11] and two pregabalin measures [12,13]). We aggregated OpenPrescribing measure performance by taking the mean percentile across all included measures for each practice or CCG in each month. Second, we calculated price-per-unit efficiency as the total identified price-per-unit savings available for each practice and CCG in each month; full methods for calculating price-per-unit savings are described elsewhere [1]. We converted this into a rate per thousand patients per month using practice population denominators. Price-per-unit efficiency is a measure of potential savings. The measured outcome in this study is the difference in price-per-unit efficiency between time periods or exposure levels. Thirdly, we calculated total spending on NHS England low-priority measures (as described elsewhere [2]) for each practice and CCG in each month. We also converted this into a rate per thousand patients per month.

We measured outcomes on a monthly basis, over a period from 3 months before the launch of the respective tool (to obtain a suitable baseline) to the most recent available data. The launch dates for the outcomes were December 2015 for the OpenPrescribing measures (ie, the OpenPrescribing service as a whole), August 2017 for the price-per-unit feature, and September 2017 for the low-priority spend feature.

Analysis

We described our analysis in detail in our prespecified analytic code, which has been shared in full [9]. Analyses were performed separately at practice and CCG levels. We used fixed-effects and between-effects panel regression in order to limit the measurement of variation within practice or CCG (ie, variation over time) and between practice and CCG, respectively.

Clinical Commissioning Group and Practice Views

In addition to the analysis prespecified in the protocol [9], we calculated summary statistics for the viewed number of practices and CCGs for each outcome in order to provide further context to the analysis.

Before and After Viewing

To measure the change in prescribing outcomes over time (within CCG or practice effects), we used a fixed-effects linear panel regression to remove the effect of time-invariant (between-practice) characteristics. We first used a univariable model and then added calendar month to the model in order to adjust for underlying national changes over time. This should leave only differences over time between practices that have and have not been viewed on OpenPrescribing.

Before Viewing

To measure differences between practices that have and have not been viewed on OpenPrescribing, we used between-effects linear panel regression. This was a simple univariable model to test the hypothesis, with the between-effects model removing any effects occurring over time. In order to remove any potential influence of OpenPrescribing, we used the 3-month period prior to the above launch dates for each outcome (as described in the outcomes section).

Results

Clinical Commissioning Group and Practice Views

Of the 207 CCGs included in the study, all were counted as exposed (≥2 views in the same month) for the mean measure outcome during at least 1 month, while 127 (61.4%) were viewed for the price-per-unit outcome and 68 (32.9%) for the low-priority prescribing outcome. We included 7318 practices in the study; of them, 4578 (62.56%) were viewed in at least 1 month for the mean measure outcome, 279 (3.81%) for the price-per-unit outcome, and 59 (0.81%) for the low-priority outcome.

Prescribing Before and After Viewing OpenPrescribing

Table 1 shows the change in prescribing outcomes measured at CCG level during and after each CCG was first viewed on OpenPrescribing. Table 2 shows the same but at practice level. Univariable results include secular trends that exist regardless of any influence of OpenPrescribing. These crude unadjusted data are presented only for reference; multivariable results account for secular trends and show the impact of OpenPrescribing views. There was no change in mean OpenPrescribing measure score at either CCG or practice level. Although there was a significant change at practice level in the univariable analysis, this effect was eliminated by adjusting for the calendar month.

Table 1. Clinical Commissioning Group (CCG)-level results of fixed-effects linear panel regression showing the change in each outcome before and after the corresponding CCG page on OpenPrescribing.net was viewed. For the measure scores, higher is worse.

Outcome/Time period	Mean (SD)	Univariable		Adjusted for calendar month	
		Change	95% CI	Change	95% CI
Mean measure score (%)					
Not viewed	50.1 (8.1)	Reference	N/A[a]	Reference	N/A
Month first viewed	50.0 (8.4)	0.00	−0.24 to 0.23	−0.09	−0.42 to 0.23
After first view	49.9 (8.5)	−0.05	−0.14 to 0.04	−0.26	−0.60 to 0.08
Price per unit (£ per 1000 patients)					
Not viewed	351.19 (165.06)	Reference	N/A	Reference	N/A
Month first viewed	335.94 (154.03)	−78.28	−92.32 to −64.24	−10.30	−21.87 to 1.26
After first view	283.72 (135.58)	−132.75	−143.45 to −122.06	−14.70	−25.56 to −3.84
Low-priority measures (£ per 1000 patients)					
Not viewed	191.68 (59.48)	Reference	N/A	Reference	N/A
Month first viewed	192.79 (62.17)	−11.13	−15.33 to −6.93	−2.13	−6.41 to 2.15
After first view	194.51 (64.40)	−9.94	−14.59 to −5.29	1.08	−3.91 to 6.07

[a]N/A: not applicable.

Table 2. Practice-level results of fixed-effects linear panel regression showing the change in each outcome before and after the corresponding practice page on OpenPrescribing.net was viewed. For the measure scores, higher is worse.

Outcome/Time period	Mean (SD)	Univariable		Adjusted for calendar month	
		Change	95% CI	Change	95% CI
Mean measure score (%)					
Not viewed	45.8 (8.6)	Reference	N/A[a]	Reference	N/A
Month first viewed	46.0 (8.8)	−0.35	−0.46 to −0.24	−0.07	−0.18 to 0.05
After first view	46.0 (8.8)	−0.52	−0.57 to −0.47	−0.04	−0.10 to 0.02
Price per unit (£ per 1000 patients)					
Not viewed	426.86 (251.99)	Reference	N/A	Reference	N/A
Month first viewed	419.51 (229.25)	−98.57	−116.96 to −80.18	−31.49	−48.12 to −14.86
After first view	385.70 (220.13)	−148.36	−163.26 to −133.46	−40.42	−54.04 to −26.81
Low-priority measures (£ per 1000 patients)					
Not viewed	185.35 (154.91)	Reference	N/A	Reference	N/A
Month first viewed	234.62 (205.06)	−12.84	−34.87 to 9.18	−4.33	−26.32 to 17.66
After first view	250.62 (223.00)	−13.15	−41.05 to 14.74	−3.11	−30.97 to 24.76

[a]N/A: not applicable.

There was a reduction in available price-per-unit savings after CCGs (Table 1) and practices (Table 2) were viewed on OpenPrescribing. The effect size was greater at practice level (−£40.42 per thousand patients per month; 95% CI −54.04 to −26.81) than at CCG level (−£14.70 per thousand patients per month; 95% CI −25.56 to −3.84). In the univariable analysis, there was a much greater effect size due to the overall trend of decreasing available savings over the study period, but the effect remained after adjustment for the calendar month.

Multiplying the estimated (per thousand patient) saving by the CCG and practice populations in the "after looking" months gives a total estimated saving of £1.47 million (95% CI £384 thousand to £2.56 million) at CCG level and £243 thousand (95% CI £162 thousand to £326 thousand) at practice level in practices and CCGs whose data were viewed. It is possible that some of these savings will overlap, and therefore, it is not appropriate to add the two figures together to create total savings. Extrapolating these savings figures to all CCGs and practices across England, if all institutions' data were viewed, it would generate an estimated annual saving of £9.7 million at CCG level and £26.8 million at practice level.

Table 3. Clinical Commissioning Group (CCG)-level results of between-effects linear panel regression, showing the differences in each outcome, before each feature was launched, between CCG pages that were subsequently viewed on OpenPrescribing.net at various levels and those that were not. For the measure scores, higher is worse.

Outcome/Number of views	CCGs, n (%)	Mean (SD)	Change	95% CI
Mean measure score (%)				
0	0 (0)	N/A[a]	N/A	N/A
32-135	69 (33.5)	50.6 (8.2)	Reference	N/A
139-228	70 (34.0)	50.4 (7.5)	−0.20	−2.93 to .52
231-2219	67 (32.5)	48.9 (8.7)	−1.74	−4.50 to 1.01
Price per unit (£ per 1000 patients)				
0	40 (19.3)	365.67 (130.69)	Reference	N/A
1-3	66 (31.9)	366.31 (154.46)	0.64	−61.16 to 62.43
4-9	47 (22.7)	416.62 (186.29)	50.95	−15.40 to 117.29
10-137	54 (26.1)	444.80 (169.25)	79.12	14.79 to 143.46
Low-priority measures (£ per 1000 patients)				
0	55 (26.6)	198.19 (52.31)	Reference	N/A
1	57 (27.5)	200.15 (59.52)	1.96	−20.40 to 24.32
2-4	47 (22.7)	193.84 (55.02)	−4.36	−27.85 to 19.14
5-52	48 (23.2)	196.76 (75.82)	−1.44	−24.80 to 21.93

[a]N/A: not applicable.

Table 4. Practice-level results of between-effects linear panel regression, showing the differences in each outcome, before each feature was launched, between practice pages that were subsequently viewed on OpenPrescribing.net at various levels and those that were not. For the measure scores, higher is worse.

Outcome/Number of views	Practices, n (%)	Mean (SD)	Change	95% CI
Mean measure score (%)				
0	381 (5.0)	45.6 (8.5)	Reference	N/A[a]
1-4	3024 (39.3)	45.7 (8.5)	0.11	−0.76 to 0.99
5-8	2013 (26.2)	46.1 (8.4)	0.49	−0.41 to 1.39
9-343	2278 (29.6)	46.7 (8.3)	1.21	0.32 to 2.10
Price per unit (£ per 1000 patients)				
0	6695 (91.6)	477.29 (259.28)	Reference	N/A
1	298 (4.1)	480.02 (262.51)	3.95	−24.62 to 32.52
2	114 (1.6)	544.99 (315.40)	69.42	23.84 to 114.99
3-49	205 (2.8)	524.52 (383.27)	47.59	13.38 to 81.81
Low-priority measures (£ per 1000 patients)				
0	6891 (94.2)	187.72 (155.65)	Reference	N/A
1	308 (4.2)	229.80 (175.17)	41.88	25.42 to 58.35
2-26	119 (1.6)	251.37 (240.24)	63.94	37.81 to 90.08
Insufficient views for additional tertile	N/A	N/A	N/A	N/A

[a]N/A: not applicable.

The total "available" savings calculated by the tool for the time after CCGs or practices were viewed were £31.3 million at CCG level and £2.4 million at practice level. In our paper [1], we estimated that around half of these "available" savings might be "achievable." This means that around 10% of the achievable savings were realized at CCG level and around 20% realized at practice level.There was no change in the total spend on low-priority measures at CCG or practice level. The small reduction seen at CCG level in the univariable analysis is again eliminated after adjustment for the calendar month.

Pre-Existing Differences Between Practices That Have or Have Not Been Viewed

Table 3 shows the differences in prescribing outcomes before each service was launched, at CCG level, stratified according to the level of views after the service was launched for each CCG. Table 4 shows the same at practice level. CCGs that have been viewed had higher available price-per-unit savings (ie, they were less cost-efficient as prescribers) but were similar for the other two outcome measures. For practices, those that have been viewed were worse for price per unit and low-priority spending but similar with respect to the mean measure score.

Discussion

Principal Findings

We found that the total available price-per-unit savings decreased following views of OpenPrescribing. This saving corresponds to a total measured decrease in spend of £243 thousand at practice level and £1.47 million at CCG level, between the feature launch and end of follow-up (August to November 2017), for practices and CCGs where the tool was viewed. We found a greater saving per patient for the practice-level exposure, but a higher overall estimated saving at CCG level, due to a greater exposed population at the CCG level. Our analytic methods make every effort to remove confounding effects, such as differences between institutions or national secular trends. Extrapolating the observed savings nationally would generate a total saving of £9.7 million per year at CCG level and £26.8 million per year at practice level. Savings from this new tool were calculated with only 1 to 3 months of follow-up data and may increase over time. We did not find any changes in the overall prescribing score or low-priority measure spend; possible reasons are discussed below. Contrary to expectations, we found that institutions whose data were viewed on OpenPrescribing were overall performing more poorly on prescribing measures prior to viewing.

Strengths and Weaknesses

We were able to remove between-practice (time-invariant) confounding effects with the use of fixed-effects linear panel regression. This meant that only differences occurring over time, before and after viewing, were measured. In addition to this, adjusting for calendar month allowed us to appropriately remove the effect of any overall national trends over time, which are independent of any effect that OpenPrescribing might have. There is a theoretical possibility of there being a systematic difference between the underlying secular trends between exposed and unexposed practices. However, for the outcome where we saw an effect (price per unit), practitioners can only identify savings by using our tool, as this is a novel method of identifying savings, requiring computationally intensive calculations setting the individual CCG or practice's prescribing in the context of all other organizations' prescribing for every drug-dose pair. This, therefore, very strongly militates against any such possibility of confounding.

By fully prespecifying our methods and publishing our protocol and analytic code in advance of conducting our analysis, we reduced, as far as possible, the impact of any potential conflict of interest. We were only able to use a proxy indicator of each institution's use of the OpenPrescribing tools, since we cannot determine who exactly is viewing the website, only that a given institution's pages have been viewed. We can identify some traffic that originates from NHS IP (Interet Protocol) addresses, and anecdotally, we receive regular feedback and queries from NHS users. However, we are not able to reliably estimate the proportion of traffic that is from NHS use. This is because not all NHS organizations use an NHS internet protocol address, and some traffic is likely to be from NHS users accessing the site from private or mobile internet connections. This is likely to have added noise to the data and, as a result, is likely to have led to our underestimating any impact from the tools, as persons not associated with a practice or CCG can view the site (meaning an institution is incorrectly counted as exposed) but cannot impact on prescribing choices. It is difficult to estimate how great this effect might be.

The number of views for each feature on the site varied substantially, largely because newer tools (such as price per unit and low-priority measures) have only been available for a short period. This means it was only possible to measure the effects on a relatively small proportion of all practices for these tools. In addition, we specified the start time for possible impact on the standard prespecified prescribing measures as December 2015 because this is when the OpenPrescribing site launched. However, very few measures were available at the initial site launch, and these were added gradually over the following 2 years. This substantially reduced our ability to detect an impact on the standard prescribing measures; however, there are no methodologically straightforward means to account for this variation in the characteristics of the exposure over time.

Findings in Context

There are many providers of services related to medicines optimization [3-6]. Such providers make varying degrees of claims, with some just being a description of the services provided, while others make strong claims of efficacy. For example, ScriptSwitch (a point-of-care tool that makes automated suggestions of preferred alternative medication options when a prescription is initiated) claims to have "delivered over £50 million in savings to the NHS" [4], but it is not obvious how this was measured or over what period these savings were made. Another example is from Oracle (a commercial database vendor whose software is used by NHS Business Services Authority to make prescribing data available to a small number of NHS users with passwords), which claims that it has "enabled antibiotic prescribing to be reduced by 7%" [14,15]. However, we are aware of no evidence being given for this very substantial claim, which must be interpreted in the context of a pre-existing downward trend in antibiotic prescribing [16], and extensive work to reduce antibiotic prescribing following the Chief Medical Officer's prioritization of the issue in 2013 [17,18]. Lastly, Prescribing Services ("Risk Stratification in Prescribing & Screening" tool) have published an attempt to measure the impact of its tool and claimed a reduction in emergency admissions; however, the methodology used is not described in any detail [19].

Policy Implications and Interpretation

We found a substantial positive impact from the use of our prescribing data tool. We also show that it is possible to conduct a robust analysis of the impact of such a tool. We will continue to monitor the effectiveness of existing and new features in order to refine the tools and monitor impact. In our view, commissioners of health services should expect robust evaluations, conducted pragmatically and at low cost at the point of care, for all such digital tools making claims for positive impact on population health and cost-effectiveness.

We found the price-per-unit savings to be higher per patient at practice level than CCG level; this is to be expected, as tailored prescribing changes from this tool are best identified at the level of individual practices, as discussed in prior work [1]. However, these savings might be achieved more simply and efficiently through a national policy change.

The lack of a positive effect on the overall prescribing measure might be explained, in part, by the construction of the mean score. There are 34 different measures making up the mean score, so improvements made by a practice or CCG focusing on one measure, in particular, is likely to be drowned out by the noise of other measures. While we would like to have measured the impact on each measure individually, this would make the analysis extremely complex due to the variety between measures and would also make interpretation more difficult. It would have been preferable to use a more parametric method to summarize the measures, such as mean Z-score, but this was not possible due to nonnormal distributions (eg, bimodal). Additionally, in contrast with the price-per-unit outcome, where OpenPrescribing is the only known source of determining such savings, it is possible that many practices and CCGs are already aware of many of the issues and had existing work to improve prescribing independently of OpenPrescribing use. This is true for many of the standard OpenPrescribing measures and for the NHS England low-priority measures, which obtained some publicity when they were announced. Other possible reasons for the lack of effect with the low-priority outcome are the potential for changes in one type of prescribing to be lost in the noise of the others and the lack of follow-up time since launch (at most 3 months).

Prior to this analysis, we hypothesized that practices or CCGs that looked at OpenPrescribing might already have better overall prescribing than those who do not, on the grounds that institutions who are proactively engaged with their data are likely also to be more effective at improving their prescribing. In fact, we found that the opposite is true for some outcomes. However, this may reflect the way the OpenPrescribing site operates, in that various features highlight practices that are performing the least well on specific prescribing measures; for example, when examining the performance on one measure for all practices in a CCG, practices are ordered from worst to best, increasing the likelihood that worse practices will be clicked on when browsing. Similarly, the OpenPrescribing “email alerts” service uses various statistical process control methods to highlight specific practices and CCGs with worse prescribing [20].

Conclusions

We found a clinically significant positive impact from the use of our prescribing data tool. We also show that it is possible to conduct a robust analysis of the impact of such a tool. We will continue to monitor the performance of the OpenPrescribing services as more follow-up time accrues and as features are added and enhanced. Our methods and full prespecification may represent a good template for similar impact assessments on services that aim to improve health care.

Acknowledgments

This work was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre Oxford, the Health Foundation (ref. 7599), the NIHR School for Primary Care Research and National Health Service England (ref. 327). The funders had no role in the conduct of the research, and were not involved in review or approval of the manuscript for publication.

Conflicts of Interest

All authors are, in whole or in part, funded by grants related to the OpenPrescribing project. As described in the paper, we have used open, pre-registered methods and analytic code, in order to reduce the impact of this conflict of interest as far as possible. BG has received funding from the Health Foundation, the National Institute for Health Research (NIHR) School of Primary Care Research, the NIHR Biomedical Research Centre Oxford, the West of England Academic Health Sciences Network and National Health Service England for work on United Kingdom prescribing data. BG has additionally received funding from the Laura and John Arnold Foundation, the Wellcome Trust and the World Health Organization to work on better use of data in medicine and receives personal income from speaking and writing for lay audiences on the misuse of science.

References

1. Croker R, Walker AJ, Bacon S, Curtis HJ, French L, Goldacre B. New mechanism to identify cost savings in English NHS prescribing: minimising 'price per unit', a cross-sectional study. BMJ Open 2018 Dec 08;8(2):e019643 [FREE Full text] [doi: 10.1136/bmjopen-2017-019643] [Medline: 29439078]
2. Walker AJ, Curtis HJ, Bacon S, Croker R, Goldacre B. Trends and variation in prescribing of low-priority treatments identified by NHS England: a cross-sectional study and interactive data tool in English primary care. J R Soc Med 2018 Jun;111(6):203-213 [FREE Full text] [doi: 10.1177/0141076818769408] [Medline: 29787684]

3. Oracle. NHS Business Services Authority Gains Better Insight into Data, Identifies circa GBP100 Million (US$156 Million) in Potential Savings in Just Three Months URL: https://www.oracle.com/uk/customers/nhsbsa-1-advanced-analytics.html [accessed 2018-04-11]
4. Optum. ScriptSwitch URL: http://www.optum.co.uk/how-we-help/scriptswitch.html [accessed 2018-04-11] [WebCite Cache ID 6ynJMiGv5]
5. Prescribing Services. Our Mission URL: https://www.prescribingservices.org/our-mission/ [accessed 2018-04-11] [WebCite Cache ID 6ynJkLNgh]
6. Prescribing Support Services. URL: http://prescribingsupportservices.co.uk/ [accessed 2018-04-11] [WebCite Cache ID 6ynNunmk4]
7. Gray J, Bevan G, Cripps M, Jani A, Ricciardi W. How to Get Better Value Healthcare. Oxford, UK: Offox Press; 2017.
8. Walker A. Open Science Framework. Measuring the impact of OpenPrescribing on real world prescribing: Registration URL: https://osf.io/9aj53/register/565fb3678c5e4a66b5582f67 [accessed 2018-09-17]
9. Walker A. Open Science Framework. 2018 Feb 12. Measuring the impact of OpenPrescribing on real world prescribing URL: https://osf.io/9aj53/ [accessed 2018-04-11] [WebCite Cache ID 6ytEBn3wY]
10. OpenPrescribing. Prescribing measures URL: https://openprescribing.net/measure/ [accessed 2017-07-31] [WebCite Cache ID 6ynOtsMcn]
11. OpenPrescribing. Direct Oral Anticoagulants (DOACs) by all CCGs URL: https://openprescribing.net/measure/doacs/ [accessed 2017-03-13] [WebCite Cache ID 6ynP1CLoJ]
12. OpenPrescribing. Prescribing of pregabalin by all CCGs URL: https://openprescribing.net/measure/pregabalin/ [accessed 2017-03-13] [WebCite Cache ID 6ynQ5ev13]
13. OpenPrescribing. Prescribing of pregabalin (total mg) by all CCGs URL: https://openprescribing.net/measure/pregabalinmg/ [accessed 2018-04-11] [WebCite Cache ID 6ynQC0BsH]
14. Oracle. PR Newswire. 2017 Oct 03. NHS Business Services Authority Helps Improve Patient Care with Oracle: Rapid, Large-scale Data Analysis Helps to Save £581 Million in Two Years URL: https://www.prnewswire.com/news-releases/nhs-business-services-authority-helps-improve-patient-care-with-oracle-300529747.html [accessed 2018-04-04] [WebCite Cache ID 6ytEHLTyq]
15. Giles-Haigh A. Digital people. 2018 Jan 31. How Data Science helped identify potential savings of over £581m for the NHS URL: https://digitalpeople.blog.gov.uk/2018/01/31/how-data-science-helped-identify-potential-savings-of-over-581m-for-the-nhs/; [accessed 2018-04-04] [WebCite Cache ID 6ytEN75xv]
16. Curtis H, Walker A, Mahtani K, Goldacre B. Time trends and geographical variation in prescribing of antibiotics in England 1998-2017. J Antimicrob Chemother 2018 Sep 15:-. [doi: 10.1093/jac/dky377] [Medline: 30239809]
17. Davies SC, Fowler T, Watson J, Livermore DM, Walker D. Annual Report of the Chief Medical Officer: infection and the rise of antimicrobial resistance. Lancet 2013 May 11;381(9878):1606-1609. [doi: 10.1016/S0140-6736(13)60604-2] [Medline: 23489756]
18. Gov.uk. 2013 Sep 10. UK 5 Year Antimicrobial Resistance Strategy 2013 to 2018 URL: https://www.gov.uk/government/publications/uk-5-year-antimicrobial-resistance-strategy-2013-to-2018 [accessed 2018-04-16] [WebCite Cache ID 6ytETNXe4]
19. Eclipse Radar. Impact Assessment 2015 URL: https://static1.squarespace.com/static/5649ce94e4b0c76d8c6eaafc/t/59aeab62e5dd5b60a9de9fd4/1504619365980/Eclipse+Radar+-+Impact+Assesment+2016.pdf
20. Walker AJ, Bacon S, Croker R, Goldacre B. Detecting change in comparison to peers in NHS prescribing data: a novel application of cumulative sum methodology. BMC Med Inform Decis Mak 2018 Jul 09;18(1):62 [FREE Full text] [doi: 10.1186/s12911-018-0642-6] [Medline: 29986693]

Abbreviations

CCG: Clinical Commissioning Group
NHS: National Health Service

Edited by G Eysenbach; submitted 04.05.18; peer-reviewed by S Speedie, A Motulsky; comments to author 01.09.18; revised version received 24.09.18; accepted 01.10.18; published 16.01.19.

Please cite as:
Walker AJ, Curtis HJ, Croker R, Bacon S, Goldacre B
Measuring the Impact of an Open Web-Based Prescribing Data Analysis Service on Clinical Practice: Cohort Study on NHS England Data
J Med Internet Res 2019;21(1):e10929
URL: http://www.jmir.org/2019/1/e10929/
doi:10.2196/10929
PMID:30664459

Original Paper

Application of Efficient Data Cleaning Using Text Clustering for Semistructured Medical Reports to Large-Scale Stool Examination Reports: Methodology Study

Hyunki Woo[1*], BS; Kyunga Kim[1,2*], PhD; KyeongMin Cha[1], MS; Jin-Young Lee[3], MD, PhD; Hansong Mun[3], MD, PhD; Soo Jin Cho[3], MD; Ji In Chung[3], MD, PhD; Jeung Hui Pyo[3], MD; Kun-Chul Lee[4], PhD; Mira Kang[1,3], MD, PhD

[1]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea
[2]Statistics and Data Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea
[3]Center for Health Promotion, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[4]Jason TG, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Mira Kang, MD, PhD
Center for Health Promotion
Samsung Medical Center
Sungkyunkwan University School of Medicine
81 Irwon-ro, Gangnam-gu
Seoul, 06351
Republic of Korea
Phone: 82 2 3410 3882
Fax: 82 2 3410 0054
Email: mira90.kang@samsung.com

Abstract

Background: Since medical research based on big data has become more common, the community's interest and effort to analyze a large amount of semistructured or unstructured text data, such as examination reports, have rapidly increased. However, these large-scale text data are often not readily applicable to analysis owing to typographical errors, inconsistencies, or data entry problems. Therefore, an efficient data cleaning process is required to ensure the veracity of such data.

Objective: In this paper, we proposed an efficient data cleaning process for large-scale medical text data, which employs text clustering methods and value-converting technique, and evaluated its performance with medical examination text data.

Methods: The proposed data cleaning process consists of text clustering and value-merging. In the text clustering step, we suggested the use of key collision and nearest neighbor methods in a complementary manner. Words (called values) in the same cluster would be expected as a correct value and its wrong representations. In the value-converting step, wrong values for each identified cluster would be converted into their correct value. We applied these data cleaning process to 574,266 stool examination reports produced for parasite analysis at Samsung Medical Center from 1995 to 2015. The performance of the proposed process was examined and compared with data cleaning processes based on a single clustering method. We used OpenRefine 2.7, an open source application that provides various text clustering methods and an efficient user interface for value-converting with common-value suggestion.

Results: A total of 1,167,104 words in stool examination reports were surveyed. In the data cleaning process, we discovered 30 correct words and 45 patterns of typographical errors and duplicates. We observed high correction rates for words with typographical errors (98.61%) and typographical error patterns (97.78%). The resulting data accuracy was nearly 100% based on the number of total words.

Conclusions: Our data cleaning process based on the combinatorial use of key collision and nearest neighbor methods provides an efficient cleaning of large-scale text data and hence improves data accuracy.

(J Med Internet Res 2019;21(1):e10013) doi:10.2196/10013

KEYWORDS

data cleaning; text clustering; key collision; nearest neighbor methods; OpenRefine

Introduction

In all of the industries, including the medical field, complex and diverse (structured, semistructured, unstructured) data have been growing dramatically for decades [1-3]. Although most health data have been digitalized, it is still not easy to handle medical records such as examination reports or physician's notes because they are historically based on paper records and generated data mainly in semistructured or unstructured forms. In addition, they may contain a variety of nonidentical duplicates, typographical errors, inconsistencies, and data entry problems [4-7].

High performance analysis requires clean and high-quality data to yield reliable results [8-11]. Therefore, efficient data cleaning takes precedence to improve the quality of data and obtain accurate analysis results [12]. However, researchers are commonly faced with many obstacles in transforming the data into a clean and high-quality dataset owing to diverse patterns of typographical errors and duplicates.

For text analysis of semistructured or unstructured data, we can use a paid program such as SAS Content Categorization (SAS Institute Inc) or IBM Watson Content Analytics (IBM) [13,14]. However, these programs are very expensive and are not readily available to individual researchers because they are mainly sold to companies or research groups. Also, these programs require extensive practice and experience.

Data cleaning using Excel's "remove duplicates" function has been done before, but it is mostly unpractical to clean the data using Excel tools. Some of the nonidentical duplicates still remain because they are not recognized as duplicates when special characters or punctuations appear [5,6,15,16]. Duplicate detection tools such as the Febrl system, TAILOR, and BigMatch were also used in cleaning data. However, Febrl has usability limitations such as slowness, unclear error messages, and complicated installations [17-20]. The listed programs are rather complex to the average users who do not have experience with programming and language functions.

Many researchers who interpret and clean the local datasets are domain experts and are not familiar with the programming language [21]. Thus, researchers need user friendly cleaning tools. OpenRefine can identify all types of strings and remove duplicates without the difficulties of programming and is a free, open source tool. OpenRefine contains the following 2 clustering methods: key collision methods and nearest neighbor methods. We proposed a data cleaning process using both text clustering methods in OpenRefine to improve accuracy of semistructured data.

Methods

We performed data cleaning of 574,266 stool examination reports conducted at Samsung Medical Center from 1995 to 2015. Data for this study were extracted from DARWIN-C, the clinical data warehouse of Samsung Medical Center. According to the data cleaning process proposed in Figure 1, we conducted data cleaning by clustering and merging parasite names and investigated its performance.

As described in Figure 1, the proposed data cleaning process consists of the following 4 steps: preprocess, text facet, systematic cleaning, and manual cleaning. In the preprocess, only names related to parasites (ie, helminth or protozoa) in raw text data were extracted using the regular expression functions of STATA MP 14.2 version [22]. The extracted words were then uploaded on OpenRefine 2.7. In the text facet step, the number of occurrences was browsed for each word.

The systematic cleaning step consists of text clustering and value-merging. Two clustering methods (ie, key collision and nearest neighbor) are used in a complementary manner to identify word clusters, each of which is expected to contain a correct word and its wrong representations with diverse forms of typographical errors (called "wrong values"). Key collision methods work by creating an alternate representation of a key that contains only the most significant or meaningful parts of a string and by clustering different strings together based on the same key. Because key collision methods are fast and simple in a variety of contexts, they have been often used for text clustering. We sequentially used 4 key collision methods including fingerprint, N-gram fingerprint, Metaphone3, and Cologne phonetic in OpenRefine. Nearest neighbor methods (also known as kNN) are widely used for clustering as well. These methods are slower but more accurate because they calculate the distance between each value. We sequentially used two nearest neighbor methods, the Levenshtein distance method and Prediction by Partial Matching method in OpenRefine. We combined both methods to enhance the accuracy [23].

For each identified cluster, the wrong values are converted to their correct word by value-merging. Because OpenRefine provides a convenient user interface that lists the correct word and its wrong values in each cluster in descending order of occurrence frequency, researchers can easily recognize the correct word and conduct the value-merging task. For "*Clonorchis sinensis*" in stool examination report data, a variety of wrong expressions were noticed in the same cluster, such as clonorchis sinesis, clnorchis sinensis, clonorchis cinensis, clonrchis sinensis, and clornorchis sinensis (Figure 2). By looking at the word list, we were able to efficiently choose "*Clonorchis sinensis*" as the correct word and make a quick decision to convert all the others to "*Clonorchis sinensis*" In the final step, the remaining words that did not belong to any cluster were investigated and manually cleaned when necessary.

Figure 1. Flow chart of our data cleaning process. PPM: prediction by partial matching.

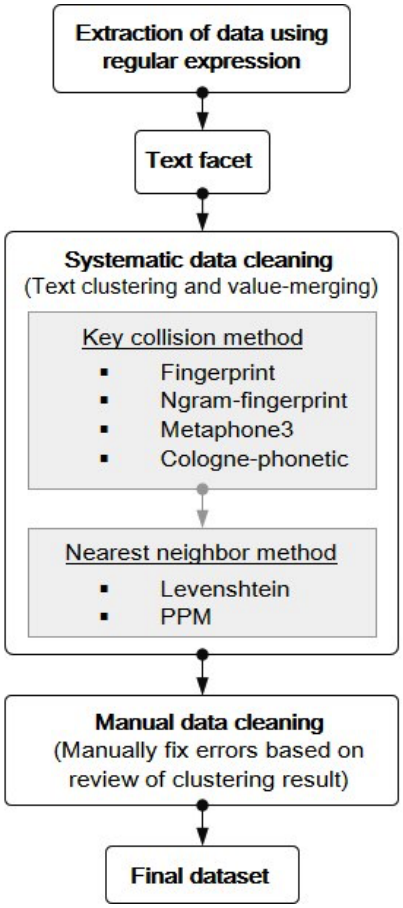
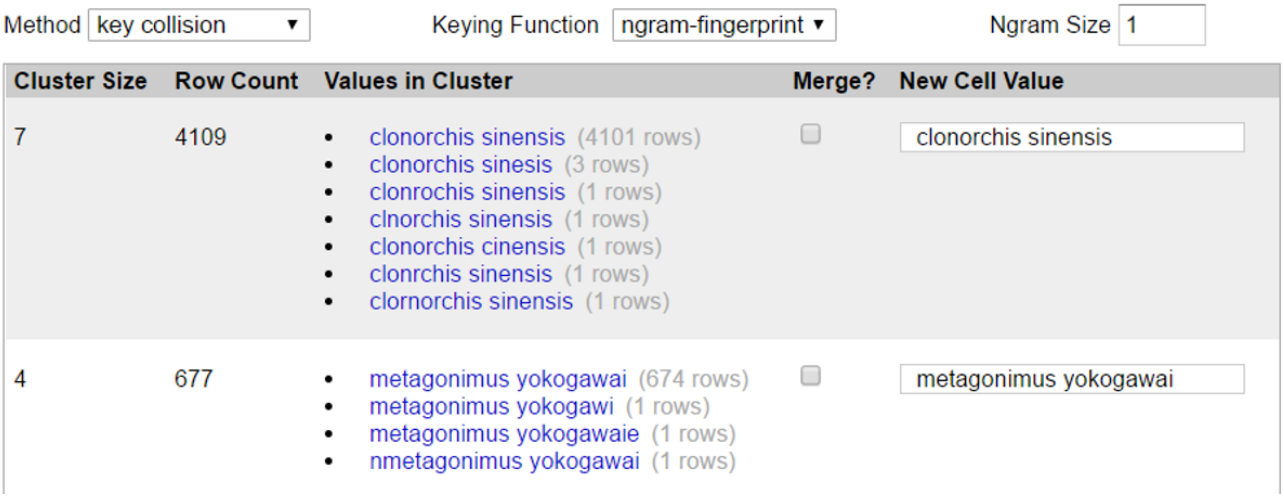


Figure 2. Representative screenshot of OpenRefine interface used for value-merging task.



Results

A total of 1,167,104 words in 574,266 stool examination reports were surveyed, and words not related to the names of helminth or protozoa were excluded from the study. We discovered 30 correct words and 45 patterns of typographical errors and duplicates (Multimedia Appendix 1). The key collision methods were able to cluster the patterns of typographical errors and duplicates with the correct word except for 6 patterns. The nearest neighbor methods were able to cluster the patterns of typographical errors and duplicates with the correct word except for 2 patterns (Table 1).

Table 1. List of typographical errors that could not be clustered with the correct word by each method.

Correct word	Typographical error	Key collision	Nearest neighbor
Negative	• Native	✗[a]	✗
	• Negaitve	✓[b]	✗
Endolimax	• Eolimax	✗	✓
	• Endolix	✗	✓
Entamoeba	• Etamoeba	✗	✓
Lamblia	• Lamdlia	✗	✓
	• G.lamblia	✗	✓

[a]✗: Typographical error is not clustered with correct word.

[b]✓: Typographical error is clustered with correct word.

Table 2. Correction rates by each method.

Method	Correction rate by the number of typographical error patterns[a], %	Correction rate by the number of typographical error words[b], %
Key collision	86.67	91.67
Nearest neighbor	95.56	97.22
Using both	97.78	98.61

[a]The number of corrected typographical error patterns divided by the total number of typographical error patterns multipled by 100 (%).

[b]The number of corrected typographical error words divided by the total number of typographical error words multipled by 100 (%).

The word “native” was the only pattern not clustered as “negative” out of all typographical errors by any clustering method because of the high inconsistency rate of the 2 words (2/6 characters, 33%). All typographical errors and duplicates except “native” were clustered correctly. We achieved a high correction rate of 98.61% by the number of typographical error words and 97.78% by the number of typographical error patterns when using both clustering methods (Table 2). After systematic data cleaning of 1,167,104 words, only 1 word with a typographical error remained and was revised manually. Thus, the accuracy of systematic data cleaning was nearly 100% based on the number of total words.

Discussion

Many researchers have made great efforts to study data analytics methodology, but there have been relatively few studies on data cleaning methodology for unexpected typographical errors [24,25]. It is rare to find a report that quantitatively analyzes the performance of data cleaning methods because they are often undocumented and used in nonofficial ways [24]. In this study, we suggested an efficient way of data cleaning for large-scale medical text data and investigated its cleaning performance. Although several methods of text analysis exist, it is not easy for general researchers to use these methods. Most methods are not readily available or have limitations in usability. Therefore, there is a need for more feasible and user friendly methods for cleaning large-scale text datasets.

We employed OpenRefine for data cleaning because of the following advantages. First, individual researchers can easily access and use OpenRefine because it is a free and open source tool. Second, OpenRefine provides researchers with an easy interface to clean the data without difficulties of programming. Third, one can easily fix rare typographical errors (which are not automatically corrected) manually and have the opportunity to modify false positive clustering [6,23].

However, we still need much effort to review each clustering result and decide whether to merge, especially in cases where the number of clustering is extremely large. In addition, formal technical support for OpenRefine is not available, and it is supported by user forums or communities. Despite these limitations, OpenRefine is a useful and effective support tool for labor-intensive and time-consuming data cleaning of semistructured data.

Our data cleaning process can be applied to other types of semistructured text data because we observed that the combinatorial use of key collision and nearest neighbor methods resulted in efficient and reliable data cleaning.

Acknowledgments

This study was supported by Samsung Medical Center grant (SMX1170601).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Patterns of parasite names in stool examination reports.

[PDF File (Adobe PDF File), 39KB - jmir_v21i1e10013_app1.pdf]

References

1. Zhang Y, Qiu M, Tsai C, Hassan MM, Alamri A. Health-CPS: Healthcare Cyber-Physical System Assisted by Cloud and Big Data. IEEE Systems Journal 2017 Mar;11(1):88-95. [doi: 10.1109/Jsyst.2015.2460747]
2. Das T, Kumar P. Big data analytics: A framework for unstructured data analysis. Int J Eng Sci Technol 2013;20135(1):A.
3. Tsai C, Lai C, Chiang M, Yang LT. Data Mining for Internet of Things: A Survey. IEEE Commun. Surv. Tutorials 2014;16(1):77-97. [doi: 10.1109/Surv.2013.103013.00206]
4. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
5. Groves A. Beyond Excel: how to start cleaning data with OpenRefine. Multimedia Information and Technology 2016;201642(2):18-22.
6. Ham K. OpenRefine (version 2.5). http://openrefine.org. Free, open-source tool for cleaning and transforming data. J Med Libr Assoc 2013 Jul;101(3):233-234 [FREE Full text] [doi: 10.3163/1536-5050.101.3.020]
7. Gallant K, Lorang E, Ramirez A. Tools for the digital humanities: a librarian's guide. 2014. URL: https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/44544/ToolsForTheDigitalHumanities.pdf?sequence=1 [accessed 2018-11-14] [WebCite Cache ID 73uqoKFfO]
8. Chu X, Ilyas IF. Qualitative data cleaning. Proc. VLDB Endow 2016 Sep 01;9(13):1605-1608. [doi: 10.14778/3007263.3007320]
9. Wang L, Jones R. Big Data Analytics for Disparate Data. American Journal of Intelligent Systems 2017;20177(2):39-46. [doi: 10.5923/j.ajis.20170702.01]
10. Zhang S, Zhang C, Yang Q. Data preparation for data mining. Appl Artif Intell 2003 May-Jun;17(5-6) 2003:375-381. [doi: 10.1080/08839510390219264]
11. Anagnostopoulos I, Zeadally S, Exposito E. Handling big data: research challenges and future directions. J Supercomput 2016 Feb 25;72(4):1494-1516. [doi: 10.1007/s11227-016-1677-z] [Medline: 26811110]
12. Rahm E, Do H. Data cleaning: Problems and current approaches. IEEE Data Eng Bull 2000;200023(4):3-13.
13. Chakraborty G, Pagolu M, Garla S. Text mininganalysis: practical methods, examples,case studies using SAS. Cary, NC: SAS Institute; 2014. URL: http://support.sas.com/publishing/pubcat/chaps/65646.pdf [accessed 2018-11-14] [WebCite Cache ID 73uraoski]
14. Zhu W, Foyle B, Gagné D, Gupta V, Magdalen J, Mundi A. IBM Watson Content Analytics: Discovering Actionable Insight from Your Content. New York, USA: IBM Redbooks; 2014.
15. Katsanevakis S, Gatto F, Zenetos A, Cardoso A. Management of Biological Invasions. 2013. How many marine aliens in Europe URL: https://pdfs.semanticscholar.org/4dbe/0bc865391bd3a6100e112e7046675341ba18.pdf [accessed 2018-11-14] [WebCite Cache ID 73uswuwIP]
16. Dallas DC, Guerrero A, Khaldi N, Borghese R, Bhandari A, Underwood MA, et al. A peptidomic analysis of human milk digestion in the infant stomach reveals protein-specific degradation patterns. J Nutr 2014 Jun;144(6):815-820 [FREE Full text] [doi: 10.3945/jn.113.185793] [Medline: 24699806]
17. Hassanien A, Azar A, Snasel V, Kacprzyk J, Abawajy J. Big Data in Complex Systems: Challenges and Opportunities Berlin, Germany. New York City, United states of America: Springer Publishing Company; 2015.
18. Selvi P, Priyaa D. A Perspective Analysis on Removal of Duplicate Records using Data Mining Techniques: A Survey. International Journal of Engineering Technology Science and Research 1;20163(12):36-41.
19. Higazy A, El TT, Yousef A, Sarhan A. Web-based Arabic/English duplicate record detection with nested blocking technique. 2013 Presented at: Computer Engineering & Systems (ICCES), 8th International Conference on IEEE; 2013; Cairo, Egypt p. 313-318.
20. Elmagarmid AK, Ipeirotis PG, Verykios VS. Duplicate Record Detection: A Survey. IEEE Trans. Knowl. Data Eng 2007 Jan;19(1):1-16. [doi: 10.1109/Tkde.2007.250581]
21. Larsson P. courses.cs.washington.edu. 2013. URL: https://courses.cs.washington.edu/courses/cse544/13sp/final-projects/p12-plarsson.pdf [accessed 2018-11-14] [WebCite Cache ID 73utiScvt]
22. Medeiros R. Using regular expressions for data management in Stata. West Coast Stata Users' Group Meetings, Stata Users Group 2007:-.
23. Clustering In Depth. 2016 URL: https://github.com/OpenRefine/OpenRefine/wiki/Clustering-In-Depth [accessed 2018-11-14] [WebCite Cache ID 73utyFlHG]

24. Maletic J, Marcus A. Data Cleansing: Beyond Integrity Analysis. IQ 2000:2000-2209.
25. Chu X, Ilyas I, Krishnan S, Wang J. Data cleaning: Overview and emerging challenges. 2016 Presented at: Proceedings of the International Conference on Management of Data. ACM; 2016; San Francisco, CA, USA p. 2201-2206.

Edited by G Eysenbach; submitted 06.02.18; peer-reviewed by W Raghupathi, S Barteit, F Shen, F Wang; comments to author 30.07.18; revised version received 23.09.18; accepted 12.10.18; published 08.01.19.

Please cite as:
Woo H, Kim K, Cha K, Lee JY, Mun H, Cho SJ, Chung JI, Pyo JH, Lee KC, Kang M
Application of Efficient Data Cleaning Using Text Clustering for Semistructured Medical Reports to Large-Scale Stool Examination Reports: Methodology Study
J Med Internet Res 2019;21(1):e10013
URL: https://www.jmir.org/2019/1/e10013/
doi:10.2196/10013
PMID:30622098

Original Paper

Use of In-Game Rewards to Motivate Daily Self-Report Compliance: Randomized Controlled Trial

Sara Taylor[1], MS; Craig Ferguson[1], MS; Fengjiao Peng[1], MS; Magdalena Schoeneich[2], MBA; Rosalind W Picard[1], ScD

[1]Affective Computing Group, Department of Media Arts and Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States
[2]Digital Accelerator, Takeda Pharmaceuticals, Cambridge, MA, United States

Corresponding Author:
Sara Taylor, MS
Affective Computing Group
Department of Media Arts and Sciences
Massachusetts Institute of Technology
77 Massachusetts Avenue
E14-348V
Cambridge, MA, 02142
United States
Phone: 1 6128608622
Email: sataylor@mit.edu

Abstract

Background: Encouraging individuals to report daily information such as unpleasant disease symptoms, daily activities and behaviors, or aspects of their physical and emotional state is difficult but necessary for many studies and clinical trials that rely on patient-reported data as primary outcomes. Use of paper diaries is the traditional method of completing daily diaries, but digital surveys are becoming the new standard because of their increased compliance; however, they still fall short of desired compliance levels.

Objective: Mobile games using in-game rewards offer the opportunity to increase compliance above the rates of digital diaries and paper diaries. We conducted a 5-week randomized control trial to compare the completion rates of a daily diary across 3 conditions: a paper-based participant-reported outcome diary (Paper PRO), an electronic-based participant-reported outcome diary (ePRO), and a novel ePRO diary with in-game rewards (Game-Motivated ePRO).

Methods: We developed a novel mobile game that is a combination of the idle and pet collection genres to reward individuals who complete a daily diary with an in-game reward. Overall, 197 individuals aged 6 to 24 years (male: 100 and female: 97) were enrolled in a 5-week study after being randomized into 1 of the 3 methods of daily diary completion. Moreover, 157 participants (male: 84 and female: 69) completed at least one diary and were subsequently included in analysis of compliance rates.

Results: We observed a significant difference ($F_{2,124}$=6.341; P=.002) in compliance to filling out daily diaries, with the Game-Motivated ePRO group having the highest compliance (mean completion 86.4%, SD 19.6%), followed by the ePRO group (mean completion 77.7%, SD 24.1%), and finally, the Paper PRO group (mean completion 70.6%, SD 23.4%). The Game-Motivated ePRO (P=.002) significantly improved compliance rates above the Paper PRO. In addition, the Game-Motivated ePRO resulted in higher compliance rates than the rates of ePRO alone (P=.09). Equally important, even though we observed significant differences in completion of daily diaries between groups, we did not observe any statistically significant differences in association between the responses to a daily mood question and study group, the average diary completion time (P=.52), or the System Usability Scale score (P=.88).

Conclusions: The Game-Motivated ePRO system encouraged individuals to complete the daily diaries above the compliance rates of the Paper PRO and ePRO without altering the participants' responses.

Trial Registration: ClinicalTrials.gov NCT03738254; http://clinicaltrials.gov/ct2/show/NCT03738254 (Archived by WebCite at http://www.webcitation.org/74T1p8u52)

(J Med Internet Res 2019;21(1):e11683) doi:10.2196/11683

KEYWORDS

self-reports; protocol compliance; recreational games

Introduction

Clinicians rely on self-reports to collect a variety of patient data (including mood or pain reports and descriptions of changing symptoms) for routine practice and during clinical trials. In many cases, patients are asked to complete paper diaries at regularly spaced prespecified times. However, it has been shown that patients often do not fill out diaries at the required time but instead fake compliance by filling them out later in batches [1]. Specifically, Stone et al found that only 11% of paper diaries were completed during the appropriate 30 min time window even though patients filled in 90% of the diaries [1].

Thus, it is critical to find a way to increase diary compliance. A common way to improve diary compliance is to transition from paper diaries to digital diaries. Even before smartphones were available, researchers were using palmtop computers to compare adherence rates between digital and paper diaries. Jamison et al had 36 participants monitor their pain daily for 1 year. They found that those participants that used digital means to record their pain completed an average of 71.5% (261/365) days, whereas those in the paper group only recorded their pain an average of 17.8% (65/365) days [2]. Palermo et al conducted a similar trial having 60 children report their pain over 7 days. They found that the digital group filled out significantly more diaries (average of 6.6/7 days or 94% completion) than the paper diary group (average of 3.8/7 days or 54% completion) [3].

With the rise of the smartphone and its near ubiquity (77% of Americans are now smartphone owners [4]), one of the simplest ways to increase diary compliance is to transition from paper diaries to digital diaries. These electronic diaries have compliance rates typically ranging from 60 to 80% over 4 weeks or less [5,6]. Experience sampling, also called ecological momentary assessment, is one of the methods used to gather an individual's experiences in real time by asking them to stop what they are doing and record their experiences [7-9]. Nevertheless, this method is highly interruptive and cannot be used for long periods without losing participant engagement unless they are highly compensated.

Recently, researchers have started exploring the possibility of using game design techniques, especially mobile games, to increase compliance to various behaviors. Played on the ubiquitous smartphone, these games have captured the attention of a wide variety of demographics and are often targeted to specific subgroups to further increase game-playing compliance. In 2015, 51.3% of mobile phone users played a mobile game at least once per month, and this rate is expected to grow to 63.7% of mobile phone users by 2020 [10]. Furthermore, 77% of teens report playing mobile games on their mobile phone or tablet [11] and 85% of children who play mobile games play at least a few times a week [12]. Successful games use common design techniques and mechanics to produce a game loop that repeatedly draws players back on a regular schedule and encourages the player to watch ads, share on social media, or pay money to get special rewards in the game.

There are many definitions of what can be construed as a game; however, 1 common theme is that a game must provide a challenge or a goal that requires skill to overcome or achieve. Typically, games also provide rewards (often called in-game rewards) that have an inherent value to help the player overcome a challenge and achieve the end goal. On the other hand, there are countless examples of attempts at including game-like features, also known as gamification, into various programs and apps in an attempt to increase engagement with the system in question. Many gamification systems use points, badges, or leaderboards to try to encourage extended engagement. We note that gamification is distinct from a game as these points and badges do not directly help the player progress toward an in-game end goal; instead, they simply mark the individuals' progress and provide no inherent value.

Some well-known examples of gamification include Fitbit [13], Apple Watch [14], Nike+ [15], Khan Academy [16], Wikipedia [17], Stack Overflow [18], Lyft [19], and many others. Although successful with some subgroups of people, these gamification systems all fail at engaging people for an extended period [20]. For example, Hanus et al included badges and leaderboards into an educational program and found that it led to lower satisfaction, lower motivation, and lower grades compared with the students who had the traditional educational program [21]. Anderson et al found that badges work when people are close to earning them, but then the activity of the user returns to baseline immediately after the badge is earned [22]. Other studies have shown that individuals are demotivated by leaderboards once they are behind [23,24]. In addition, Koivisto et al found that although everyone tires of simple gamification techniques, younger children tire faster than others, making the addition of true game design principles especially important in younger populations [25]. However, when a full game is developed and important game design principles such as fantasy, challenge, and curiosity [26] are implemented into a system, they become more engaging than their *gamification* counterparts.

Therefore, we hypothesize that using in-game rewards instead of gamification techniques will increase the diary completion compliance above simply transitioning to digital diaries. To test this hypothesis, we conducted a 5-week user study to compare the completion rates of a daily diary across 3 conditions: a paper diary, a digital diary, and digital diary with in-game rewards. Furthermore, as there are legitimate concerns that a game may influence how the participant answers the questions in the diary [27], we will also test the difference in the answer distributions between the different study groups.

Methods

Overview

In the following sections, we describe our novel game called "The Guardians" that was designed to increase compliance to a daily task. We also describe our experimental protocol and statistical methods.

The Guardians: A Diary System With In-Game Rewards

To test our hypothesis that in-game rewards can encourage engagement over a long period, we designed and developed a fully functional mobile game called "The Guardians" (see Figures 1 and 2) that was used to collect daily self-reported data.

To develop The Guardians, our team first conducted a series of pilot studies with potential users. We prototyped several different versions including several pen-and paper-based versions and simple text-based versions of the game to test the basic game mechanics. We conducted in-person interviews with children and adults (N=14; aged 7-50 years) and beta tested early prototypes to learn which game mechanics and art styles were most appealing. The final version of The Guardians used in this study was developed using Unity (Unity Technologies ApS), built for both the Android and iOS platforms, and incorporated the feedback learned from these early studies.

"The Guardians" is a mix of 2 popular game genres: idle games (a game that does not require constant player input to progress in the game but often progresses exponentially in response to a few user inputs [28], eg, Cookie Clicker and Adventure Capitalist) and pet collection games (a game where the player seeks to collect all of the pets in a set, eg, Pokemon or NekoAtsume). Besides being very popular, these genres are known for their ability to engage populations for long periods and for their appeal to a wide age range.

At the beginning of the game, the player is introduced to the Guardian, who asks for help in collecting Light to push back Darkness from the land. The player is given a pet and told that pets generate Light and is encouraged to care for them. Each day, the player has the opportunity to answer a set of diary questions and receive a reward, namely a new pet or an upgrade to an existing pet. Importantly, the player is reminded each time they start a diary that their diary responses do not affect their reward. This is reinforced by showing the silhouette of the reward before the player starts the diary, showing that it has not changed during the process (see Figure 1). This reminder is intended to encourage players to be honest in their answers instead of trying to get some response from the game. The only data returned by the diary to the game is a binary flag indicating that the diary has been completed.

The diary is accessed through the blue button in the top left of Figure 1. When the player is connected to the internet, they can click this button to open a diary. First, the silhouette of the reward is shown, and then a webview is opened showing the diary questions. When the player completes the diary, their answers are time-stamped and securely stored on an outside server, and the player is presented with their in-game reward. Importantly, the players do not have to fill out the diary to access the game nor do they have to play the game to access the diary. Thus, we have architected the system to try to avoid having the reward influence the content that people might enter into the diary while trying to simultaneously influence how often they go to fill out the diary.

After a diary entry is filled out and the player receives a pet as a reward, the pet helps the player generate Light for the Guardian, even when the app is closed. The player can customize, rename, and interact with the pets. Spending Light to level up the pets can increase the power of a pet. Players can also solve a simple placement puzzle each day for an additional bonus (see Figure 2 for screenshots of the leveling and placement puzzle menus). Along the way, the player completes several quests by reaching various levels for the pets. Once all of the quests are completed, the Guardian is awakened and Darkness is defeated. We designed "The Guardians" to be played in small periods of time (approximately 2-5 min each day) so that players would not tire quickly.

The version of the game used in this study was designed with enough content to engage the player for approximately 35 days, whereas the platform built to host the game is extensible for providing significantly longer experiments in the future and for enabling customized diary questions. The Guardians game was also designed to be accessible to all skill levels and to provide motivational content for a broad range of ages. Thus, 6-year-olds just beginning to read can enjoy receiving pets and interacting with them, whereas other optional aspects of the game, such as the positioning and leveling puzzles, are designed to be interesting and challenging for adults.

Experimental Setup

Although Guardians was designed to increase compliance across a wide age range, the main focus of this paper is on individuals aged 6 to 24 years because this is the specific age group being targeted in an upcoming clinical trial. Before this new platform is used in a true clinical trial, we need to validate if the Game-Motivated electronic-based participant-reported outcome diary (ePRO) could improve daily adherence while not impacting the content of the responses. Thus, individuals aged 6 to 24 years were recruited via posters and internet postings to participate in a 5-week study about mood and mobile game habits in return for receiving a US $100 Amazon gift card at the end of the study. Recruitment was conducted from October 9, 2017, to November 15, 2017.

Interested individuals were screened for their age and ownership of a smartphone (only Android and iPhone owners were eligible for this study). Eligible individuals, and their parent/guardian if the individual was under 18 years, were emailed consent forms, which informed them of the study protocol, namely, that the study would consist of 3 parts:

1. A Web-based prestudy survey (approximately 30 min long) about demographics, general mood, and mobile game habits.
2. A diary on mood and mobile game habits (6 questions total) to be filled out daily for 5 weeks (35 days)—individuals were informed that they would be asked by study personnel to complete these diaries with an app or on paper.
3. A Web-based poststudy survey (approximately 20 min long) about general mood, the usability of the daily diary system, and qualitative feedback about the study.

Figure 1. Screenshots of The Guardians while the player completes a daily diary. Players collect pets by filling out a daily diary. The pets are needed to complete quests and gather Light to awaken the guardian.

Figure 2. Screenshots of leveling menu and housing placement puzzle.

Individuals were informed that they needed to complete the prestudy survey and at least four of the first seven diaries to be eligible for the first US $40 and that they could receive the remaining US $60 if they filled out the poststudy survey. No other requirements were made on filling out daily diaries beyond the minimum of 4 in the first week. Individuals who were asked to complete the daily diaries on paper were given an additional US $10 to cover the cost of mailing the diaries back to the study personnel.

Immediately after providing consent, but before completing the prestudy survey, participants were randomized without regard to age or gender into 1 of 3 groups for the daily diaries: a paper-based participant-reported outcome diary (Paper PRO), an ePRO, or our novel ePRO diary with in-game rewards (Game-Motivated ePRO). The 6 daily multiple-choice questions were identical for all 3 groups. Specifically, participants were asked to classify their mood, report on how much time they spent playing mobile games and how long they spent outside in the past 24 hours, how many ads they watched for in-game rewards, and what the weather (temperature and precipitation) was like in the past 24 hours (see Multimedia Appendix 1 for the exact survey content). Any diary that did not have all 6 questions completed was discarded. These questions were selected to study relationships between mood, mobile games, and the weather; however, this analysis is beyond the scope of this paper. Nevertheless, we note that it is common to ask individuals to answer short multiple-choice questions daily, and the mood question is of particular interest for health-related daily reports as it might be part of a clinical trial [29-31].

The paper PRO group also received 1 additional question. This question was used to confirm when the diary was completed and required viewing a Web page that displayed a word that changed every hour. Any paper diary with mismatched “word-of-the-hour” and recorded date were marked as incomplete. We note that 55% (23/42) participants in the Paper PRO group returned at least one diary with a mismatched “word-of-the-hour” and recorded date.

The ePRO group’s app was a simple native iOS or Android app that displayed a webview with the daily diary form. The Game-Motivated ePRO included the exact same webview and form. Therefore, the ePRO and Game-Motivated ePRO groups had identical daily diaries; however, the Game-Motivated ePRO participants received an in-game reward (ie, a pet) when they completed a daily diary, whereas the ePRO participants were just shown a screen thanking them for completing the diary and reminding them to return the next day. Participants in the Game-Motivated ePRO group were informed that they did not need to play the game to fill out their daily diary.

All participants were asked to complete the diary immediately after waking, but the diaries were available to complete from 3:30 am to 3:30 am each day (localized to the participant’s time zone). This time was chosen after learning that most college-aged individuals went to sleep between midnight and 3 am; thus, the diaries were reset while the majority of participants would be sleeping. Participants in the ePRO and Game-Motivated ePRO groups received a notification from their respective apps if they had not completed the diary by 9:30 am local time.

Participants were not informed that the main goal of the study was to measure daily diary compliance over the course of the study. Instead, they were told that the study was about mood, mobile game habits, and the weather because the daily questions referred to these topics. The entire study was done remotely and participants only had email contact with the study personnel.

The Massachusetts Institute of Technology (MIT) Committee on the Use of Humans as Experimental Subjects approved this protocol (MIT IRB Protocol #1708061907).

Statistical Methods

Participants

We compared the individuals who were randomized but did not start the study and those who were randomized and did start the study (ie, filled out a prestudy survey) to see whether there were any age differences. As participants were blind to which study group they were in until after completing the prestudy survey, we hypothesized that there would be no significant differences found. Specifically, we conducted a 2-way analysis of variance (ANOVA) test to examine the interaction between choosing to participate and study arm on age. Tukey post hoc tests were conducted when statistically significant differences were found.

We hypothesized that there would be some portion of participants in each study arm that completed the prestudy survey, but did not complete any of the daily diaries. Specifically, we hypothesized that the Paper PRO group would have more participants that failed to complete a single daily diary than either the ePRO group or the Game-Motivated ePRO group as completing and returning diaries on paper presents a much larger barrier than completing diaries digitally. However, we did not expect a significant difference between the ePRO and Game-Motivated ePRO groups. To conduct this hypothesis test, we used the N-1 chi-square test in a pairwise manner to determine whether the study arms had a significantly different proportion of participants completing at least one daily diary. We adjusted for multiple comparisons by using the Bonferroni technique. We also computed the 95% CIs by using the adjusted Wald technique. Other than these tests, all of our other analyses, described below, focus on data from participants who completed at least one diary.

Game Engagement Analysis

To determine if participants in the Game-Motivated ePRO group actually engaged with the game during the study, we computed 4 measures of whether an individual engaged with the main components of the game (leveling pets, completing the placement puzzle, and interacting with the pets). Specifically, we measured (1) the average number of levels purchased daily, (2) the average number of daily housing placement changes, (3) the percentage of pets with a custom name, and (4) the average daily number of pets with an equipped cosmetic item.

To determine if higher levels of engagement correlated with the number of diaries completed, we analyzed the correlation between the number of diaries completed and the 4 engagement measures defined above. We also computed the correlation between age and these same engagement measures. In both correlation analyses, we adjusted for multiple comparisons by using the Bonferroni technique.

We are also interested in comparing the number of individuals engaged in the various components of the game; thus, we used the following thresholds to convert these measures into binary values. Individuals who were engaged with the leveling component of the game were expected to purchase at least 20 levels per pet per day as the exponential income players earn easily allows them to reach this amount. Furthermore, the minimum expected number of housing changes to optimally solve the housing placement puzzle was approximately 0.8 changes per day (some days required no changes). Therefore, we used an average of 0.5 housing placement changes per day (ie, one change every other day) as the engagement threshold for the housing puzzle. Finally, we used 50% as the engagement threshold for both the percentage of pets with a custom name and the average daily number of pets with a cosmetic item. Using these binary thresholds, we counted the number of participants who engaged with each component. We also computed the average diary completion for the combinations of engagement behaviors.

Daily Diary Completion

Beyond considering differences in the 3 study groups, we also examined age and gender differences on completion rates. Participants were split into 5 different age groups before any analyses were conducted. Specifically, the age groups were defined as children younger than 10 years—the average age of first ownership of a smartphone (ie, 6-9 years [32]), three 3-year age groups of preteens and teenagers (10-12 years, 13-15 years, and 16-18 years), and a group of young adults (19-24 years). Furthermore, some research has shown that there is a gender difference in how individuals engage with games [33,34]; therefore, we also examine gender effects in our analysis.

In some studies in which patient reported outcomes are collected, any data collected are useful regardless of which user it came from, but the more days of data collected the better. To determine the difference in average completion rates between the 3 study arms, we used a three-way ANOVA (also called a factorial ANOVA) test to consider how the diary method, age group, and gender influenced the adherence rate. Tukey post hoc tests are conducted when statistically significant differences are found.

On the other hand, some studies require high levels of completion to include an individual in the final analysis. If an individual falls below the required completion rate, study coordinators will often remove that participant from the study. Therefore, we compare the number of participants that are able to complete at least 90% of the daily diaries (ie, completing at least 32 of the 35 daily diaries). Specifically, we estimate the survival (ie, missing 3 or fewer diaries) of a participant in the study by computing the Kaplan-Meier curve [35] for each study group using the *survival* R package [36]. Then using the log-rank test [37], we can compare the survival curves of the 3 study groups. Specifically, it tests the hypothesis that at least 1 of the groups has a different survival than another against the null hypothesis that all of the groups have the same survival. If a significant *P* value is found, we will conduct a pairwise post

hoc test to determine which pairs of study groups are significantly different from each other when adjusting for multiple comparisons using the Holm method [38]. Finally, we also compare the resulting number of participants with 90% or greater completion rates in each arm of the study using pairwise N-1 chi-square tests to check for significant differences in the participants that achieved 90% completion rate.

Self-Reported Response Distributions

As mentioned previously, the daily diaries contained the same 6 multiple-choice questions. A potential concern of the game is that it could affect the person's state and cause them to respond differently to the questions [27]. We hypothesized that the method of completing the diary would not affect the answers to the mood, outside duration, number of ads watched, or the weather questions. Specifically, we hypothesized that the probability of selecting a particular answer would not differ between study arms after controlling for individual differences.

We used a Bayesian multinomial mixed effects logistic regression model in R, using the *brms* package [39], to estimate the association between the study arm and answer response. To account for clustering in the responses because of repeated measures for each individual, we included a random intercept grouped by the participant. We then computed point and interval (95% credibility) estimates of the predicted probability of selecting a particular answer response for each of the 3 study arms. If the credibility intervals of 2 study arms overlap for a given answer response, we conclude that no difference is detectable between the study arms. We complete this comparison for all pairs of study groups for each answer response. As this analysis is conducted in the Bayesian framework, we do not need to adjust for multiple comparisons.

We also performed the same analysis on the question about how long participants spent playing mobile games in the past 24 hours. We hypothesized that there would be a significant difference between the answers of the Game-Motivate ePRO group and the other groups because these individuals would be playing the Guardians regularly, especially in the 0 and 1 to 15 min responses as the game is designed to be played in approximately 5 min each day.

Average Duration of Daily Diary and System Usability Scale Scores

For the ePRO and Game-Motivated ePRO groups, we were able to track how long the participants took to complete each daily diary. Also, during the poststudy survey, participants in the ePRO and Game-Motivated ePRO evaluated the daily diary software using the System Usability Scale (SUS) [40]. As the diaries for these 2 groups were identical, we hypothesized that there would be no significant difference in the average duration or in the average SUS score between the 2 groups even when controlling for age and gender interactions using a three-way ANOVA test for each outcome measure. As before, Tukey post hoc tests were conducted when statistically significant differences were found.

Qualitative Poststudy Survey Results

Recall that at the end of the 35-day study, participants were given a Web-based poststudy survey to complete. Besides containing the SUS survey, participants were asked about what they thought the best and worst parts of the study were. Participants responded to these questions using an open format text response. These text responses were grouped into common themes and are presented in the results.

Results

Participants

Of the 372 individuals screened for the study, 309 individuals were eligible to participate and 232 individuals provided informed consent and were randomized into the 3 study groups. Of these, 197 completed the prestudy survey to begin the study: 67 participants (male: 31 and female: 36) in the Paper PRO group, 66 (male: 33 and female: 33) in the ePRO group, and 64 (male: 36 and female: 28) in the Game-Motivated ePRO group (see Figure 3 for the flow of participants in the study and Figure 4 and Table 1 for the distribution of study conditions in the different age groups).

The two-way ANOVA (Multimedia Appendix 1) did not reveal a statistically significant difference in age in the interaction between study group and participation in the study ($F_{2,226}$=0.513; P=.60), as hypothesized. Nevertheless, we did observe a significant difference in age between study groups ($F_{2,226}$=3.007; P=.05) and between individuals who completed the prestudy survey and those who did not ($F_{1,226}$=3.184; P=.08). A Tukey post hoc test revealed that individuals randomized to the paper PRO group were slightly older (mean age difference 2.06 years; 95% CI 0.03-4.09 years; P_{adj}=.05) and individuals who completed the prestudy survey were also slightly older (mean age difference 1.73 years; 95% CI −0.20 to 3.66 years; P_{adj}=.08).

A total of 44 participants never completed a single daily diary after completing the prestudy survey. There was no significant difference (P_{adj}=.57, 95% CI −20.2% to 32.9%) between the number of participants who never completed a diary in the ePRO group and Game-Motivated ePRO group; however, as hypothesized, there was a significant difference between the Paper PRO group and ePRO and Game-Motivated ePRO groups (P_{adj}<.001; 95% CI 12.2%-39.7% and P_{adj}=.06; 95% CI 3.0%-32.9%, respectively).

Game Analysis

We found that the average number of levels purchased and the average number of pets with a cosmetic item were positively correlated with the number of diaries a participant completed (P_{adj}<.001 and .02, respectively). The percentage of pets with custom names was found to be negatively correlated with the age of the participant (P_{adj}=.09) and the average number of housing changes was found to be positively correlated with the age of the participant (P_{adj}=.03, see Multimedia Appendix 1 for complete results).

Figure 3. Flow of participants through the study. ePRO: electronic-based participant-reported outcome; PRO: participant-reported outcome.

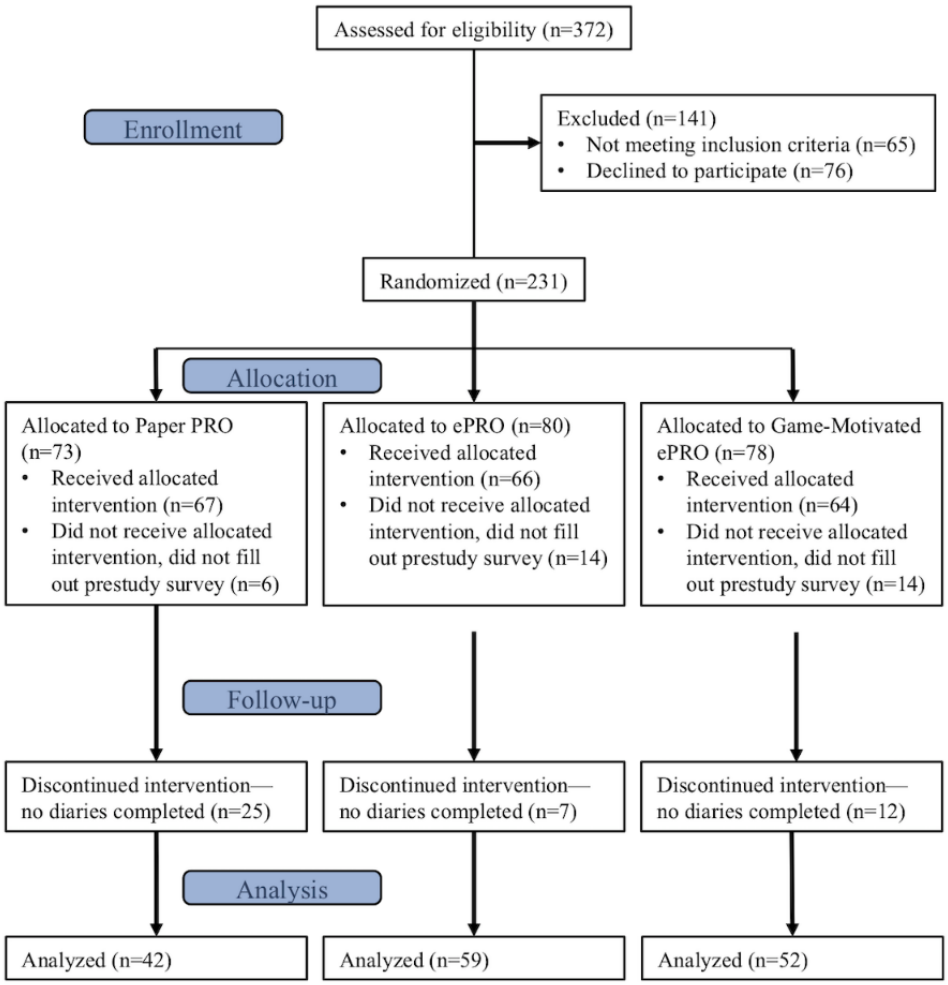


Figure 4. Distribution of participants who filled out the prestudy survey by age and daily diary condition. ePRO: electronic-based participant-reported outcome; PRO: participant-reported outcome.

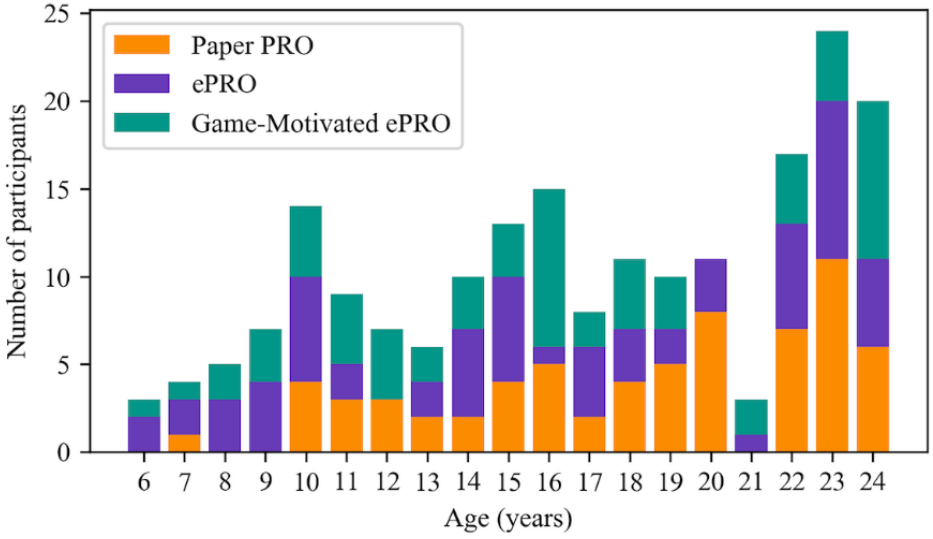


Table 1. Demographic details of participants who completed at least one daily diary by study group.

Characteristic	PRO[a] (n=42), n (%)	ePRO[b] (n=59), n (%)	Game-Motivated ePRO (n=52), n (%)
Gender			
Male	24 (57)	29 (49)	31 (60)
Female	18 (43)	30 (51)	21 (40)
Age (years)			
6 to 9	1 (2)	9 (15)	5 (10)
10 to 12	10 (24)	7 (12)	8 (15)
13 to 15	6 (14)	12 (20)	6 (12)
16 to 18	5 (12)	7 (12)	14 (27)
19 to 24	20 (48)	24 (41)	19 (37)

[a]PRO: participant-reported outcome.

[b]ePRO: electronic-based participant-reported outcome.

Table 2. Average diary completion per cluster of engagement (N=52).

Number of components engaged	n (%)	Daily diary completion, mean (SD)
All (n=4)	16 (31)	0.957 (0.051)
Three components	22 (42)	0.847 (0.198)
Two components	7 (14)	0.820 (0.141)
One component	5 (10)	0.657 (0.380)
None	2 (4)	0.971 (0.040)

We found the majority of participants in the Game-Motivated ePRO group surpassed the binary thresholds to be considered *engaged* with the leveling component (48/52, 92%), equipping their pets with cosmetic items (43/52, 83%) and solving the housing placement puzzle (39/52, 75%). However, approximately one-third engaged in personalizing their pets with custom names (n=19, 37%). Looking at the combinations of engagement patterns (see Table 2 and Figure 5), we found that 16 participants engaged with all components of the game (16/52, 31%); these participants had very high daily diary completion rates (fraction completed: mean 0.957, SD 0.051). In addition, 2 of the participants did not engage with any components of the game; however, they still had very high daily diary completion rates (fraction completed: mean 0.971, SD 0.040).

Daily Diary Completion

The Game-Motivated ePRO group had the highest compliance (mean completion 86.4%, SD 19.6%), followed by the ePRO group (mean completion 77.7%, SD 24.1%), and finally, the Paper PRO group (mean completion 70.6%, SD 23.4%).

The three-way ANOVA (see Figure 6 and Multimedia Appendix 1) revealed a statistically significant difference in completion rates between the study groups ($F_{2,124}$=6.341; P=.002) and a statistically significant interaction between study group and age group ($F_{8,124}$=2.530; P=.01). A Tukey post hoc test revealed that the completion rate of the Game-Motivated ePRO was significantly higher than the Paper PRO (mean difference 15.8%; 95% CI 5.2%-26.4%; P_{adj}=.002) and higher than the ePRO (mean difference 8.7%; 95% CI 1.0%-18.4%; P_{adj}=.09). However, there was no statistically significant difference between the ePRO and Paper PRO groups (mean difference 7.1%; 95% CI −3.2% to 17.4%; P_{adj}=.24). Therefore, our hypothesis that the Game-Motivated ePRO group would have significantly higher completion rates than both the ePRO group and the Paper PRO group was held. However, the hypothesis that the ePRO group would have a significantly higher completion rate than the Paper PRO did not hold.

The survival analysis estimated the Kaplan-Meier curves for each study group (see Figure 7) and found that the study groups were significantly different than each other (P<.001). A post hoc log rank test found that the Game-Motivated ePRO group had a significantly higher survival curve than the Paper PRO group (P_{adj}<.001) and the ePRO group (P_{adj}=.01). The ePRO group had a significantly higher survival curve than the Paper PRO group (P_{adj}=.005).

Furthermore, as shown in Figure 8, there are significant differences in the final percentage of participants completing 90% of daily diaries between each of the study arms, with the Game-Motivated ePRO having the highest percentage, followed by the ePRO and Paper PRO, with all comparisons in the hypothesized directions having P<.02. We note that the increase in the percentage of participants with high compliance in the ePRO compared with the Paper PRO (CI 7.1%-39.8%) is almost identical to the increase in the percentage of participants with high compliance in the Game-Motivated ePRO compared with the ePRO (CI 3.9%-39.7%).

Figure 5. Game-motivated electronic participant-reported outcome participants' engagement with various components of the game and the corresponding diary completion.

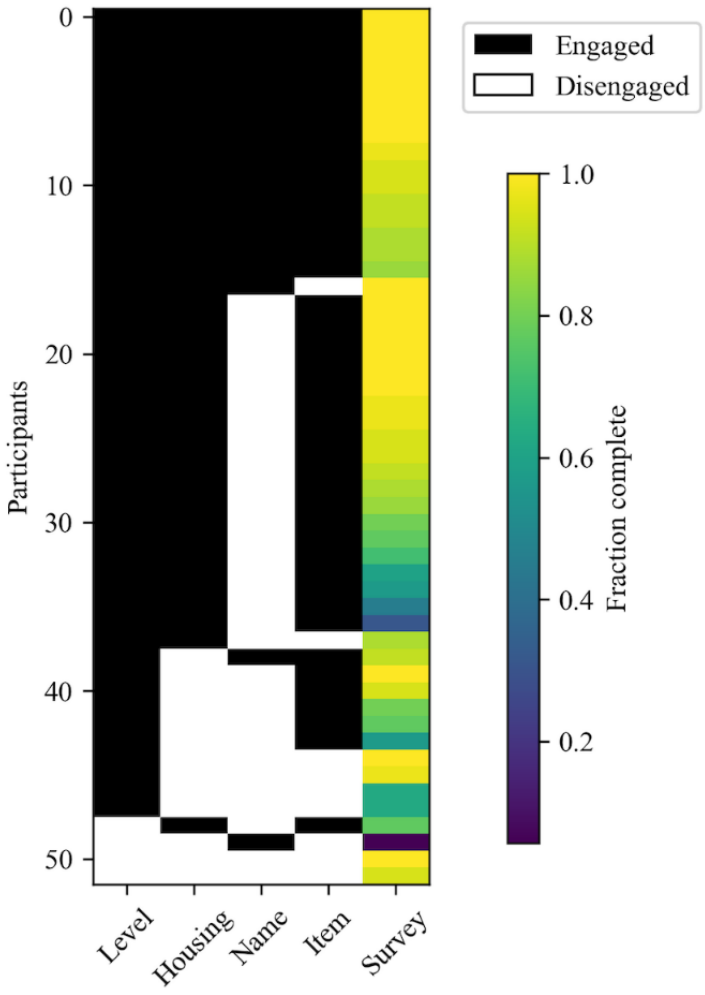


Figure 6. Average daily diary completion rate for each study arm. The error bars show the 95% CI and P values are from the Tukey posthoc test and are adjusted for multiple comparisons. ePRO: electronic-based participant-reported outcome; PRO: participant-reported outcome.

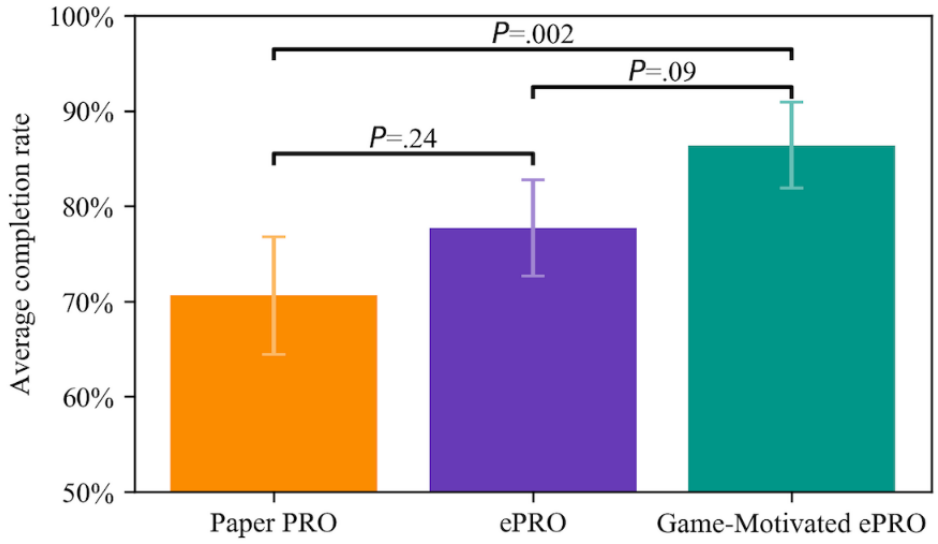


Figure 7. Kaplan-Meier curves of survival (ie, missing 3 or fewer diaries) for each study group. ePRO: electronic-based participant-reported outcome; PRO: participant-reported outcome.

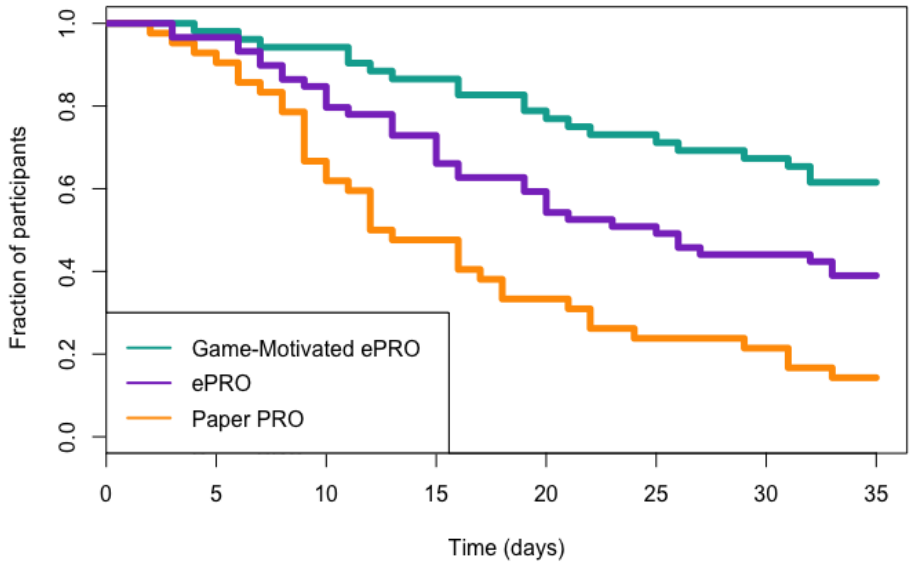
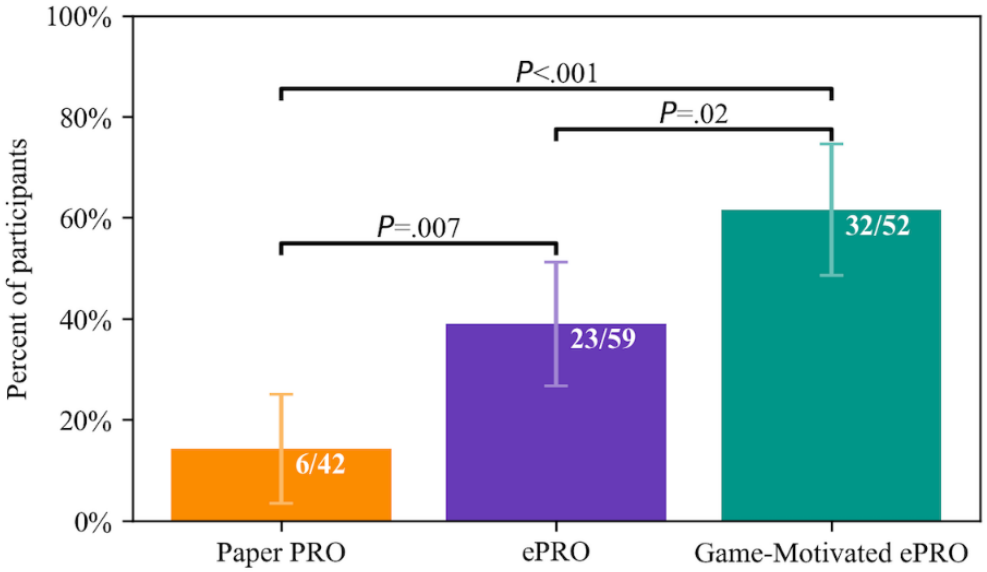


Figure 8. Percentage of participants who completed at least 90% of the daily diaries (ie, 32 out of 35 diaries) in each study arm. The 95% CI and P values are shown. ePRO: electronic-based participant-reported outcome; PRO: participant-reported outcomes.



Self-Reported Response Distributions

The results of the Bayesian multinomial mixed effects logistic regression models can be found in Multimedia Appendix 1. As all pairs of 95% credible intervals overlap for each answer response for the mood, number of ads watched, and weather questions, we conclude that no difference in the study arm has been detected for the answer selected. Therefore, there is no evidence that the probability of an answer response differs between the 3 study arms for these questions, as hypothesized. However, for the outside duration question, we note the Paper PRO and ePRO 95% credible intervals do not overlap for the "0 Minutes" response (Paper PRO: 3.3%-7.2% and ePRO: 7.3%-12.4%). Thus, we conclude that a statistically significant difference has been found in how these 2 study groups respond to this question of the daily survey.

Furthermore, we found a statistically significant difference between the ePRO and Game-Motivated ePRO groups in the probability of selecting "0 minutes" for the game duration question, as hypothesized. Specifically, we note that the 95% credibility interval for the probability of selecting "0 minutes" in the ePRO group (33.5%-49.7%) was significantly higher than the Game-Motivated ePRO group (13.0%-24.4%). We did not, however, see a significant difference in the "1-15 minute" response.

Average Duration of Daily Diary

Our hypothesis that the average duration of time spent filling out the daily diary was not significantly different between the ePRO (mean duration 41.2 seconds, SD 19.3 seconds) and Game-Motivated ePRO (mean duration 44.1 seconds, SD 33.9 seconds) study groups was also confirmed ($F_{1,91}$=0.423; P=.52). However, we did find some age and gender effects. The

three-way ANOVA (see Multimedia Appendix 1) revealed a statistically significant difference in a participants' average time spent filling out the daily diary between the age groups ($F_{4,91}$=3.705; P=.008) and between the 2 genders ($F_{1,91}$=4.256; P=.04). A Tukey post hoc test revealed that the 19- to 24-year-old participants' average duration was significantly faster than the 16- to 18-year-old participants (mean difference 20.5 seconds; 95% CI 3.1-37.9 seconds; P_{adj}=.01) and faster than the 10- to 12-year-old participants (mean difference 18.0 seconds; 95% CI −1.5 to 37.6 seconds; P_{adj}=.09). Furthermore, male participants were found to be significantly faster than female participants (mean difference 9.1 seconds; 95% CI 0.2-18.0 seconds; P_{adj}=.04).

System Usability Scale Scores

We found that the usability of the diary, measured via SUS, was high for both the ePRO (mean 86.5, SD 14.5) and Game-Motivated ePRO (mean 86.9, SD 12.6). The three-way ANOVA (see Multimedia Appendix 1) revealed a significant difference between age groups ($F_{4,88}$=2.221; P=.07) and between gender ($F_{1,88}$=4.898; P=.03). A Tukey honestly significant difference test revealed that the only statistically significant (P_{adj}<.1) difference in SUS score between age groups was between the participants aged 19 to 24 years and the 6- to 9-year-old participants, with the older participants rating the diary as more usable (mean difference 10.2; 95% CI −1.0 to 21.5; P_{adj}=.09). In addition, we found that female participants rated the diary higher than male participants (mean difference 5.2; 95% CI 0.5-10.0; P_{adj}=.03). Therefore, there was no significant difference in the SUS scores between the ePRO and Game-Motivated ePRO groups, as hypothesized.

Qualitative Results From the Poststudy Survey

Over half of the participants (84/158) said the simplicity of the daily diaries and the opportunity it gave them to reflect on their day was the best part of the study. For example, 1 participant said "being able to reflect on the previous 24 hours" was the best part of the study. Another participant reported his or her favorite part was "each survey did not take long." Moreover, 37 out of 56 (66%) of the Game-Motivated ePRO participants said the game was the best part. For example, one participant said their favorite part of the study was that "It was connected to a fun game, and it reminded me to take surveys."

In total, 29.7% of participants (47/158) said remembering to do the daily diary was the worst part of the study. Nearly 40% of these participants (18/47) came from the Paper PRO group, with several participants suggesting that the diaries be converted to digital diaries that could be completed on a smartphone. For example, one participant suggested, "Instead of paper based this could have been mobile based."

Discussion

Principal Findings

This study aimed to observe if there was a difference in daily diary completion in children and young adults between 3 different methods of diary, namely traditional paper-based diaries, digital diaries, and our custom game-motivated digital diary. As hypothesized, we observed significant differences in compliance to filling out daily diaries, with the Game-Motivated ePRO group having the highest compliance, followed by the ePRO group, and finally, the Paper PRO group. Nearly all individuals in the Game-Motivated ePRO group actually engaged with the various game mechanics, even though no such engagement was required to fill out the diaries. Importantly, there were no statistically significant differences in the content of the responses to 5 of the 6 daily questions, the average diary completion time, or in the SUS score between the ePRO and Game-Motivated ePRO groups. As mentioned previously, we expected the game duration question to have a significant difference in responses as the Game-Motivated ePRO group received a game to play during the study. Therefore, we conclude that the Game-Motivated ePRO method encouraged individuals to complete significantly more diaries without significantly altering the content of their responses.

Other researchers have also found promise in using in-game rewards to motivate individuals to complete a task. Cechanowicz et al added true game design techniques to the theme of a game show in their market research survey, including timers and points to add challenge and graphics to add to the game show theme. They found a significant difference in engagement between the full game and traditional survey and between the full game and the partial game (ie, when the challenge and graphical design elements were removed) [41]. Li et al created a full game to teach first time AutoCAD users how to use the software [42]. Participants who received the game-based version of the tutorial performed tests faster and reported higher subjective engagement levels than their traditional tutorial counterparts. Although their design is yet to be evaluated for engagement, Bindoff et al proposed a game design to help smokers regularly engage with smoking cessation content to earn currency in an idle world building game [43].

Some might worry that introducing a mobile game to increase compliance will introduce a potential *addiction* to the game that negatively influences the individual more than the benefit of reporting their symptoms. Whether a gaming addiction is officially recognized as a disorder or not, the game used in this study was designed to take only a few minutes to play and did not appear to have the negative effects one would expect from an addiction (eg, lack of control in playing the game, playing the game for excessive amounts of time, or being unable to stop playing even after negative consequences). For example, participants in the Game-Motivated ePRO group reported significantly lower rates of playing for 0 minutes a day than the other ePRO group but no differences for higher amounts of time. Thus, the game used in this study did not cause individuals to spend excessively more time playing games than the other groups. Nevertheless, future work in using games to motivate compliance should always monitor for the risks of potentially addictive behaviors.

Limitations

Our study relied on having a significant compensation: US $40 to fill out 4 diaries in the first 7 days and an additional US $60 to complete a poststudy survey 35 days later. Ideally, using a

daily diary system with in-game rewards would require little to no monetary compensation so that collecting data could be scaled to a massive number of participants without the need for large project budgets. Our "up to US $100" compensation might also account for the reason why the participants in the Paper PRO group had much higher completion rates than other studies (eg, mean completion rate of 70.6% in our study compared with 18% in the study by Jamison et al [2] and 54% in the study by Palermo et al [3]). Nevertheless, the compensation schedule in our study was designed to get individuals to start the habit of filling out the daily diaries and avoided offering compensation for each diary completed (a model used by many other daily diary studies).

As our study required individuals to own a smartphone, the proportion of the youngest participants was small in comparison with the proportion of young adults. Therefore, our results may not generalize well to individuals who rely on a parent or guardian owning a smartphone to complete an ePRO.

Although our study showed no difference in how the participants answer the questions in the daily diary, whether or not it was accompanied by the game, these results may not generalize to more complex diaries beyond the short, multiple-choice questions used in this study. Thus, future work should monitor how a game influences data collected in response to more complex diaries.

In addition, although we seek to capture long-term engagement in self-reporting compliance, this study was limited to a 5-week duration. Although this duration was longer than most daily diary studies, we have now extended our custom platform to provide content to engage individuals for longer periods of time in future studies, with the goal of understanding how engagement changes over months and through different periods of life.

Finally, although we posited that using a diary system as a part of a full game that provides in-game rewards will drive long-term engagement in a way that is not possible with simpler gamification techniques, a limitation of this study is that we do not know which elements of the game are, in fact, driving the better compliance. The novel full-narrative game has not been compared directly with other lesser gamification techniques. Future work should examine and deconstruct the various elements of the game to understand which elements contribute to increased motivation and compliance for daily diary reports.

Conclusions

Self-reports are critical to research and clinical care; however, they require persistence and motivation to complete at regular intervals, especially if they involve answering daily diary questions about unexciting or possibly even unpleasant topics (such as those which are part of many clinical trials). Although moving away from traditional paper-based surveys to mobile digital surveys has shown increased compliance over traditional paper-based ones, and this finding was replicated here, our study shows that the compliance levels of today's electronic diaries can still be improved.

We have shown that mobile game techniques, when properly implemented, can increase compliance in daily patient-reported outcomes. We have shown a significant increase in compliance over both a paper diary and a digital diary. We have shown data to support this hypothesis in a pediatric and young adult population. Furthermore, we have shown that the in-game rewards did not impact the content of the answers provided in the diaries.

Future work should seek to replicate these results when no monetary compensation is offered, in specific clinical patient populations to make sure that the difference in survey compliance rates holds, and for a period longer than 5 weeks. In addition, we intend to study the effects of engagement in older adult populations. Looking beyond patient reported outcomes and other surveys or self-reports, we also see potential for this motivation through games approach to be extended to other areas of health and well-being.

Acknowledgments

The authors thank all of the participants for joining this study. They also thank the members of the Takeda Digital Accelerator Team for their helpful comments and suggestions, especially Bruno Villetelle, Eric Perakslis, and Gina Rey. Statistical support was provided by data science specialist Steven Worthington at the Institute for Quantitative Social Science, Harvard University. Millennium Pharmaceuticals and the MIT Media Lab Consortium financially supported this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary material including daily survey questions and extended results.

[PDF File (Adobe PDF File), 498KB - jmir_v21i1e11683_app1.pdf]

Multimedia Appendix 2

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 628KB - jmir_v21i1e11683_app2.pdf]

References

1. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [doi: 10.1016/S0197-2456(02)00320-3]
2. Jamison RN, Raymond SA, Levine JG, Slawsby EA, Nedeljkovic SS, Katz NP. Electronic diaries for monitoring chronic pain: 1-year validation study. Pain 2001;91(3):277-285. [doi: 10.1016/S0304-3959(00)00450-4]
3. Palermo TM, Valenzuela D, Stork PP. A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. Pain 2004;107(3):213-219. [doi: 10.1016/j.pain.2003.10.005]
4. Pew Research Center. 2018 Feb 05. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-12-10] [WebCite Cache ID 6yExPU0V9]
5. Harris C, Daniels K, Briner R. A daily diary study of goals and affective well-being at work. J Occup Organ Psychol 2003;76:401-410. [doi: 10.1348/096317903769647256]
6. Totterdell P, Kellett S, Teuchmann K, Briner R. Evidence of mood linkage in work groups. Journal of Personality and Social Psychology 1998;74(6):1504-1515 [FREE Full text] [doi: 10.1037//0022-3514.74.6.1504]
7. Csikszentmihalyi M, Larson R, Prescott S. The ecology of adolescent activity and experience. J Youth Adolesc 1977 Sep;6(3):281-294. [doi: 10.1007/BF02138940] [Medline: 24408457]
8. Hernandez J, McDuff D, Infante C, Maes P, Quigley K, Picard R. Wearable ESM: Differences in the experience sampling method across wearable devices. In: Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services. 2016 Sep Presented at: Mobile HCI'16; September 06-09, 2016; Florence, Italy p. 195-205 URL: https://dl.acm.org/citation.cfm?doid=2935334.2935340 [doi: 10.1145/2935334.2935340]
9. Intille S, Haynes C, Maniar D, Ponnada A, Manjourides J. µEMA: Microinteraction-based Ecological Momentary Assessment (EMA) Using a Smartwatch. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Sep Presented at: UbiComp'16; September 12-16, 2016; Heidelberg, Germany p. 1124-1128 URL: http://europepmc.org/abstract/MED/30238088 [doi: 10.1145/2971648.2971717]
10. Statista. Mobile Phone Gaming Penetration in The United States from 2011 to 2020 URL: https://www.statista.com/statistics/234649/percentage-of-us-population-that-play-mobile-games/ [accessed 2018-10-03] [WebCite Cache ID 72td0S3Iz]
11. Statista. Share of Teenagers Who Play Games on Their Mobile Phone or Tablet in The United States from Spring 2013 to Fall 2016 URL: https://www.statista.com/statistics/749973/us-teens-playing-mobile-or-tablet-games/ [accessed 2018-10-03] [WebCite Cache ID 72tdMMW7h]
12. Statista. Frequency with Which Children in The United States Play Their Favorite Gaming Apps in 2015 URL: https://www.statista.com/statistics/694723/frequency-children-us-play-favorite-gaming-apps/ [accessed 2018-10-03] [WebCite Cache ID 72tdcpLvU]
13. Kosecki D. Fitbit. Presenting the Official List of Fitbit Badges - How Many Do You Have? URL: https://blog.fitbit.com/fitbit-badges/ [accessed 2018-03-27] [WebCite Cache ID 6yExiy29p]
14. Apple. Earn Activity achievements using your Apple Watch URL: https://support.apple.com/en-us/HT205406 [accessed 2018-03-27] [WebCite Cache ID 6yExnUzAd]
15. Nike. Nike+ Running Club App: What's New URL: https://www.nike.com/us/en_us/c/nike-plus/running-app-gps/overview [accessed 2018-03-27] [WebCite Cache ID 6yExt01CB]
16. Khan Academy. Badges URL: https://www.khanacademy.org/badges [accessed 2018-03-27] [WebCite Cache ID 6yExyoTX0]
17. Wikipedia. Wikipedia: Service awards URL: https://en.wikipedia.org/wiki/Wikipedia:Service_awards [accessed 2018-03-27] [WebCite Cache ID 6yEy2m8Dv]
18. Stack Overflow. Badges URL: https://stackoverflow.com/help/badges [accessed 2018-03-27] [WebCite Cache ID 6yEy71lnP]
19. Lyft Blog. Lyft Badge Glossary URL: https://blog.lyft.com/badge-glossary/ [accessed 2018-03-27] [WebCite Cache ID 6yEyS7I9Z]
20. Hamari J, Koivisto J, Sarsa H. Does Gamification Work? -- A Literature Review of Empirical Studies on Gamification. In: Proceedings of the 2014 47th Hawaii International Conference on System Sciences. 2014 Mar 10 Presented at: HICSS'14; January 06-09, 2014; Waikoloa, HI, USA p. 3025-3034. [doi: 10.1109/HICSS.2014.377]
21. Hanus MD, Fox J. Assessing the effects of gamification in the classroom: a longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. Comput Educ 2015 Jan;80:152-161. [doi: 10.1016/j.compedu.2014.08.019]
22. Anderson A, Huttenlocher D, Kleinberg J, Leskovec J. Steering user behavior with badges. In: Proceedings of the 22nd international conference on World Wide Web.: ACM; 2013 May Presented at: WWW'13; May 13-17, 2013; Rio de Janeiro, Brazil p. 95-105. [doi: 10.1145/2488388.2488398]
23. Mejia J. Grand Valley State University. 2013 Apr. Impact of Gamification and Shared Situated Displays on Smartphone Application Engagement URL: https://scholarworks.gvsu.edu/cistechlib/152 [accessed 2018-12-10] [WebCite Cache ID 74YyA5MrM]
24. Farzan R, DiMicco J, Millen D, Dugan C, Geyer W, Brownholtz E. Results from deploying a participation incentive mechanism within the enterprise. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2018 Apr Presented at: CHI'08; April 05-10, 2008; Florence, Italy p. 563-572. [doi: 10.1145/1357054.1357145]

25. Koivisto J, Hamari J. Demographic differences in perceived benefits from gamification. Comput Human Behav 2014 Jun;35:179-188. [doi: 10.1016/j.chb.2014.03.007]
26. Malone T. Heuristics for designing enjoyable user interfaces: Lessons from computer games. In: Proceedings of the 1982 Conference on Human Factors in Computing Systems. 1982 Presented at: CHI'82; March 15-17, 1982; Gaithersburg, Maryland, USA p. 63-68. [doi: 10.1145/800049.801756]
27. Lu AS, Baranowski J, Islam N, Baranowski T. How to engage children in self-administered dietary assessment programmes. J Hum Nutr Diet 2014 Jan;27(Suppl 1):5-9 [FREE Full text] [doi: 10.1111/j.1365-277X.2012.01258.x] [Medline: 22594618]
28. Alharthi S, Olaa A, Toups Z, Tanenbaum J, Hammer J. Playing to Wait: A Taxonomy of Idle Games. In: Proceedings of the 2018 Conference on Human Factors in Computing Systems. 2018 Apr Presented at: CHI'18; April 21-26, 2018; Montreal QC, Canada p. 1-15. [doi: 10.1145/3173574.3174195]
29. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. 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 Jun 08;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
30. Suhara Y, Yinzhan X, Pentland A. DeepMood: Forecasting depressed mood based on self-reported histories via recurrent neural networks. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Apr Presented at: WWW'17; April 03-07, 2017; Perth, Australia p. 715-724. [doi: 10.1145/3038912.3052676]
31. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Sep Presented at: UbiComp'15; September 07-11, 2015; Osaka, Japan p. 1293-1304. [doi: 10.1145/2750858.2805845]
32. Influence Central. Kids & Tech: The Evolution of Today's Digital Natives URL: http://influence-central.com/kids-tech-the-evolution-of-todays-digital-natives/ [accessed 2018-10-02] [WebCite Cache ID 72sNXb1if]
33. Ogletree SM, Drake R. College students' video game participation and perceptions: gender differences and implications. Sex Roles 2007 Mar 21;56(7-8):537-542. [doi: 10.1007/s11199-007-9193-5]
34. Wright J, Huston A, Vandewater E, Bickham D, Scantlin R, Kotler J, et al. American children's use of electronic media in 1997: a national survey. J Appl Dev Psychol 2001 Jan;22(1):31-47. [doi: 10.1016/S0193-3973(00)00064-2]
35. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc 1958 Jun;53(282):457. [doi: 10.2307/2281868]
36. Therneau T. The Comprehensive R Archive Network. 2015. A Package for Survival Analysis in R version 2.38 URL: https://CRAN.R-project.org/package=survival [accessed 2018-12-10] [WebCite Cache ID 73jQ7HQQz]
37. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep 1966 Mar;50(3):163-170. [Medline: 5910392]
38. Holm S. A simple sequentially rejective multiple test procedure. Scand Stat Theory Appl 1979;6(2):65-70 [FREE Full text]
39. Bürkner P. brms: an R package for Bayesian multilevel models using Stan. J Stat Soft 2017;80(1):-. [doi: 10.18637/jss.v080.i01]
40. Brooke J. SUS-A quick and dirty usability scale. In: Jordan P, Thomas B, McClelland I, Weerdmeester B, editors. Usability Evaluation in Industry. London: Taylor & Francis, Inc; 1996:189-194.
41. Cechanowicz J, Gutwin C, Brownell B, Goodfellow L. Effects of gamification on participation and data quality in a real-world market research domain. In: Proceedings of the First International Conference on Gameful Design, Research, and Applications. 2013 Oct Presented at: Gamification'13; October 02-04, 2013; Toronto, Ontario, Canada p. 58-65. [doi: 10.1145/2583008.2583016]
42. Li W, Grossman T, Fitzmaurice G. GamiCAD: a gamified tutorial system for first time autocad users. In: Proceedings of the 25th annual ACM symposium on User interface software and technology. 2012 Oct Presented at: UIST'12; October 07-10, 2012; Cambridge, Massachusetts, USA p. 103-112. [doi: 10.1145/2380116.2380131]
43. Bindoff I, de Salas K, Peterson G, Ling T, Lewis I, Wells L, et al. Quittr: the design of a video game to support smoking cessation. JMIR Serious Games 2016 Dec 01;4(2):e19 [FREE Full text] [doi: 10.2196/games.6258] [Medline: 27908844]

Abbreviations

ANOVA: analysis of variance
ePRO: electronic-based participant-reported outcome
MIT: Massachusetts Institute of Technology
PRO: participant-reported outcome
SUS: System Usability Scale

Edited by G Eysenbach; submitted 25.07.18; peer-reviewed by T Baranowski, I Bindoff, K Bul, S Amantini, R Alkoudmani; comments to author 16.08.18; revised version received 31.10.18; accepted 22.11.18; published 03.01.19.

Please cite as:
Taylor S, Ferguson C, Peng F, Schoeneich M, Picard RW
Use of In-Game Rewards to Motivate Daily Self-Report Compliance: Randomized Controlled Trial
J Med Internet Res 2019;21(1):e11683
URL: https://www.jmir.org/2019/1/e11683/
doi:10.2196/11683
PMID:30609986

Original Paper

Measurement of Quality of Life in Patients with Mycosis Fungoides/Sézary Syndrome Cutaneous T-Cell Lymphoma: Development of an Electronic Instrument

Stacey McCaffrey[1], PhD; Ryan A Black[2], PhD; Mitchell Nagao[3], PharmD; Marjan Sepassi[3], PharmD; Gaurav Sharma[3], PharmD; Susan Thornton[4], BBA; Youn H Kim[5], MD; Julia Braverman[1], PhD

[1]PatientsLikeMe, Inc, Cambridge, MA, United States

[2]Nova Southeastern University, Fort Lauderdale, FL, United States

[3]Actelion (US), Inc, South San Francisco, CA, United States

[4]Cutaneous Lymphoma Foundation, Birmingham, MI, United States

[5]Stanford University, Stanford, CA, United States

Corresponding Author:
Stacey McCaffrey, PhD
PatientsLikeMe, Inc
160 Second Street
Cambridge, MA, 02142
United States
Phone: 1 6174994003
Fax: 1 8668506240
Email: research@patientslikeme.com

Abstract

Background: Although the quality of life (QoL) plays an important role in treatment decision making and clinical management of mycosis fungoides (MF) or Sézary syndrome (SS) subtypes of cutaneous T-cell lymphomas (MF/SS-CTCLs), an MF- or SS-specific measure of QoL does not exist.

Objective: The objective of this research was to develop and validate the first QoL instrument for MF/SS-CTCL using a patient-centered approach.

Methods: A conceptual framework for the MF/SS-CTCL QoL was developed through a literature review and interviews with key opinion leaders. Concept elicitation with patients was utilized to refine the conceptual model and generate preliminary items. The items were then revised based on qualitative and quantitative feedback obtained through cognitive debriefing surveys and interviews with patients. Next, participants (N=126) completed the preliminary MF/SS-CTCL QoL and a comparator measure of health-related QoL (Skindex-29) through the PatientsLikeMe Open Research Exchange. The MF/SS-CTCL QoL was completed again 5 days later by 66 participants for the purposes of evaluating test-retest reliability. The MF/SS-CTCL QoL was finalized based on results from an empirical evaluation, which included both classical and modern test theory approaches. Specifically, this included evaluation of (1) the optimal item response theory measurement model; (2) item fit; (3) unidimensionality; (4) rating scale performance; (5) reliability; (6) test information (precision); (7) person-to-item map; (8) convergent and discriminant validity; and (9) presence of bias via differential item function.

Results: Results from the comprehensive psychometric evaluation utilizing a Rasch-Grouped Rating Scale model yielded a final 12-item instrument. The rating scale functioned as expected, and the instrument exhibited adequate person reliability (.87), good to excellent test-retest reliability (r=.89, P<.001), high levels of measurement precision, and good person-to-item targeting. The correlation between the MF/SS-CTCL QoL and the Skindex-29 (r=.852, P<.001) was significantly greater than the correlation between the MF/SS-CTCL QoL and syndrome stage (r=.260, P<.001), providing support for convergent and discriminant validity. Items did not show significant bias based on gender, age, or race. Rasch scores were converted to scaled scores with qualitative descriptive categories for ease of interpretation.

Conclusions: Empirical evaluation demonstrated strong evidence of excellent psychometric properties. Utilizing a patient-centered measure development approach ensures that this QoL instrument captures the information that is most meaningful and clinically relevant to patients.

(J Med Internet Res 2019;21(1):e11302) doi:10.2196/11302

KEYWORDS

quality of life; Rasch; patient-reported outcome; cutaneous T-cell lymphoma; mycosis fungoides; Sézary syndrome

Introduction

Mycosis fungoides (MF) and its leukemic variant Sézary syndrome (SS) represent approximately 65% of the cases of cutaneous T-cell lymphoma (CTCL), a class of non-Hodgkin's lymphomas with a relapsing course over the span of decades [1,2]. For patients with MF or SS subtypes of CTCL (MF/SS-CTCL), quality of life (QoL) plays an important role in treatment decision making and clinical management of the disease. Currently, several cancer-specific (European Organisation for Research and Treatment of Cancer tools [3] and Functional Assessment of Cancer Therapy-General [4]) and skin-specific (ItchyQoL [5], Dermatology Life Quality Index [6], Skindex-29 [7], and itch visual analog scale [8]) health-related QoL instruments exist. Although clinicians often use these instruments or a combination of these instruments to estimate QoL for patients with MF/SS-CTCL, this strategy can be time consuming and burdensome for patients. These patient-reported outcome (PRO) instruments were not specifically designed to capture the unique experiences of patients living with MF/SS-CTCL. They may contain content that is irrelevant for patients with MF/SS-CTCL and fail to capture the aspects of QoL that are most meaningful for these patients, decreasing the instrument's sensitivity in detecting important changes in QoL [9]. Therefore, a patient-centered, disease-specific PRO to measure QoL for patients with MF/SS-CTCL is urgently needed to improve the quality of care for these patients and to progress research within this clinical arena. The purpose of this study is to fill this critical gap by developing and validating the first QoL instrument specifically developed by and designed for patients with MF/SS-CTCL—the "MF/SS-CTCL QoL." The MF/SS-CTCL QoL was developed in two broad phases: (1) instrument development and (2) psychometric evaluation. Methods and results of each phase are described separately below.

Methods

Instrument Development

The purpose of instrument development was to create items that comprehensively capture the different facets of QoL that are impacted by MF/SS-CTCL. Patients were closely involved in the item development process to ensure that the final instrument evaluated aspects of QoL that are most relevant and meaningful for them. Instrument development involved three primary steps: (1) creating a conceptual framework; (2) concept elicitation; and (3) cognitive debriefing. This research was approved by the New England Institutional Review Board.

Creating the Conceptual Framework

The conceptual framework of QoL for MF/SS-CTCL patients was developed through a literature review and interviews with key opinion leaders (KOLs). Physicians and experts in the field of cutaneous lymphomas (N=3) participated in interviews to gather information related to treatment, challenges in caring for and treating patients with MF/SS-CTCL, main concerns expressed by patients, the impact of the condition on patients' well-being and daily functioning, use and availability of PRO instruments, and unmet needs within the research and patient care field.

Results from the literature review and KOL interviews highlighted the importance of evaluating condition-specific facets of QoL, such as physical functioning, emotional functioning, and social functioning. KOLs also indicated that two facets of QoL—coping and self-management—were absent from existing PRO measures and may be important for patients with MF/SS-CTCL. More information about KOL input and the conceptual model is available upon request.

Concept Elicitation

The purpose of concept elicitation was to gather patient feedback regarding their experience of living with MF/SS-CTCL and generate preliminary items. Data for concept elicitation were collected from patients through a survey conducted via the PatientsLikeMe Web-based research platform (Open Research Exchange, ORE) and follow-up interviews conducted using phone or videoconferencing. Patients were eligible to participate if they were members of PatientsLikeMe, were adults, and reported a diagnosis of MF or SS. Survey content included demographic and clinical items as well as open-ended questions pertaining to health-related QoL derived from the conceptual model. Follow-up interviews consisted of semistructured questions based on participants' responses to the survey.

The data were coded independently by two trained raters using MAXQDA software (VERBI). The coders (a research scientist and research assistant), experienced in qualitative research and coding, were trained by the senior author (JB). The raters coded the interviews independently, and any discrepancies in codes were resolved by the senior author. The codebook was finalized after a satisfactory interrater agreement (Cohen kappa of 0.65 or greater) was reached. Content saturation was assessed across patients with a saturation table where saturation was reached when no new information was obtained through data collection [10]. The codes with best agreement and highest frequencies were selected and then grouped by themes to generate the initial items for the QoL instrument.

Results From Concept Elicitation

The Web-based survey was completed by 21 participants, and 10 of those participants completed a follow-up interview. The sample comprised 67% (14/21) females, all of them being white and non-Hispanic, and the average age was 55 (SD 12.39) years. Of the 21 participants, 16 (76%) reported a diagnosis of MF, 3 (14%) reported a diagnosis of SS, and 2 (10%) did not report a diagnosis. The average disease length was 10 (SD 9.50) years, with a range of <1 year to 31 years. Among all, 14 participants reported their stage of diagnosis and indicated that they had

stage IA (8/21, 38%), stage IB (4/21, 19%), or stage IIB (2/21, 10%) disease.

Based on the qualitative analysis, 43 of the 60 codes developed from the coding scheme reached agreement of a Cohen kappa at or above 0.65. Saturation was reached after 15 patients, suggesting an adequate sample size. Thematic content analysis identified six major code groupings (treatment, impact on daily activities, emotional, social, coping and management, and symptoms and symptom burden), which were used to generate the 31-item preliminary version of the MF/SS-CTCL QoL (more details about the survey and results are available upon request).

Cognitive Debriefing

Using the same Web-based research platform (ORE) and participant inclusion criteria from the concept elicitation phase, the preliminary version of the MF/SS-CTCL QoL was administered to a sample of participants. Although a partnership between PatientsLikeMe and the Cutaneous Lymphoma Foundation (a patient-advocacy group) was made to help with patient recruitment, there was substantial overlap in participants across stages of the research study due to difficulty recruiting patients with these rare diseases. Participants were asked to complete the preliminary items and to provide specific quantitative and qualitative feedback regarding clarity or semantic ambiguity and understanding, relevance, and adequacy of each item and the response options.

Results From Cognitive Debriefing

Overall, 42 participants took part in cognitive debriefing. Of the 41 participants who chose to provide demographic information, approximately half were men (51%, 21/41) with an average age of 62 (SD 14.13; range 31-101) years. The majority (85%, 35/41) reported a diagnosis of MF. Their disease stage ranged from IA to IVA, with most participants reporting stage IA (39%, 16/41) or stage IB (17%, 7/41) disease.

Based on quantitative and qualitative cognitive debriefing results, changes were made to the instrument; items were removed and revised to improve clarity, a response option was added for patients who were in remission, and the recall period was changed from 7 days to 4 weeks. At this stage, the MF/SS-CTCL QoL contained 14 items.

Psychometric Evaluation

Participants

Patients were eligible to participate if they were members of the Web-based community (PatientsLikeMe), adults (aged ≥18 years), and reported a diagnosis of MF or SS. Participants were recruited through the PatientsLikeMe Web-based community with support from the Cutaneous Lymphoma Foundation.

Data Collection

Following consent, eligible participants completed a demographic survey, the MF/SS-CTCL QoL, and the Skindex-29 [7] through the Web-based research platform. Additionally, participants were asked to complete a second administration of the MF/SS-CTCL QoL 5 days later to evaluate the stability of item functioning.

Measures

The Skindex-29 [7] is a commonly used and valid 29-item self-report measure that evaluates health-related QoL. Specifically, the Skindex-29 covers facets of QoL such as emotional functioning, physical functioning, and symptoms with a 4-week recall period. The preliminary 14-item version of the MF/SS-CTCL QoL required patients to rate their impairment in health-related QoL over the last 4 weeks using a 1 (not at all or never) to 5 (very much or always) Likert-type rating scale. Furthermore, 4 of the items included a sixth response option, “Does not apply (I don’t have symptoms right now).”

As part of this study, participants also completed a brief demographics survey asking them to provide information about their sex, age, race, ethnicity, diagnosis, and stage of their diagnosis.

Psychometric Evaluation Procedures

The empirical evaluation included determining (1) the optimal item response theory (IRT) measurement model; (2) item fit; (3) unidimensionality; (4) rating scale performance; (5) reliability; (6) test information (precision); (7) person-to-item map; (8) convergent and discriminant validity; and (9) presence of bias via differential item function (DIF). Analyses were performed in SPSS version 24 (IBM Corp) and Winsteps version 3.74.0 (Winsteps.com).

Results

Participants

A total of 126 patients completed the survey, and 52.4% (66/126) patients completed the second administration of the survey. Most participants were non-Hispanic (115/126, 91.3%) individuals and identified as white or Caucasian (108/126, 85.7%) and most were females (74/126, 58.7%). Participants ranged in age from 22 to 86 years, with an average age of 59 (SD 13.5) years. Of the 126 participants, 118 (93.7%) reported a diagnosis of MF and 8 (6.3%) reported a diagnosis of SS. Participants indicated that they had stage IA (56/126, 44.4%), stage IB (24/126, 19.0%), or stage II or above (22/126, 17.5%) disease and 19.0% (24/126) did not know or report their stage. The average disease length was 8 (SD 7.6) years, with a range of <1 year to 35 years. A summary of participants’ prescribed treatments is presented in Table 1.

Item Descriptive Statistics

Item descriptive statistics are presented in Table 2. Responses of “Does not apply (I don’t have symptoms right now)” (score=0) were marked as missing and removed from analyses when calculating mean and SD.

Table 1. Participants' self-reported prescribed treatments.

Prescribed treatment	Value, n (%)[a]
Topical corticosteroids	79 (62.7)
Other prescribed topical treatments	40 (31.7)
Phototherapy (psoralen and ultraviolet A, ultraviolet B)	12 (9.5)
Total-skin electron beam therapy	16 (12.7)
Local radiation therapy	14 (11.1)
Oral treatments or chemotherapy	113 (89.7)

[a]Treatment % is greater than 100% due to multiple selections being allowed.

Table 2. Item descriptive statistics.

Item	Minimum score	Maximum score	Mean (SD)
1. In the past 4 weeks, how much did you worry that your mycosis fungoides or Sézary syndrome may get worse?	1	5	2.74 (1.26)
2. In the past 4 weeks, how often did you feel hopeless because of having mycosis fungoides or Sézary syndrome?	1	5	2.01 (1.13)
3. In the past 4 weeks, how frustrated were you by the unpredictability of mycosis fungoides or Sézary syndrome?	1	5	2.67 (1.36)
4. In the past 4 weeks, how often did you feel depressed or sad because of mycosis fungoides or Sézary syndrome?	1	5	2.15 (1.03)
5. In the past 4 weeks, how confident did you feel about managing your mycosis fungoides or Sézary syndrome?	1	5	2.94 (1.11)
6. In the past 4 weeks, to what extent were you able to cope with the daily demands (symptom impact and management, treatment, side effects, appointments, etc) of mycosis fungoides or Sézary syndrome?	1	5	3.47 (1.27)
7. In the past 4 weeks, how severe were your mycosis fungoides or Sézary syndrome symptoms?	1	5	1.95 (1.05)
8. In the past 4 weeks, how burdensome was your mycosis fungoides or Sézary syndrome treatment?	1	5	2.20 (1.02)
9. In the past 4 weeks, how much did your mycosis fungoides or Sézary syndrome limit your daily activities (work inside and outside of the house, self-care such as cooking, cleaning, getting dressed, etc)?	1	5	1.79 (1.25)
10. In the past 4 weeks, how much did mycosis fungoides or Sézary syndrome limit your ability to wear clothes you wanted to?	1	5	2.28 (1.49)
11. In the past 4 weeks, how often did mycosis fungoides or Sézary syndrome (the condition or associated treatment) leave you too tired to work or do daily activities?	1	5	2.11 (1.20)
12. In the past 4 weeks, how much did mycosis fungoides or Sézary syndrome negatively affect your relationships with others close to you?	1	5	1.73 (1.10)
13. In the past 4 weeks, how often did you feel that others do not understand what you are going through with mycosis fungoides or Sézary syndrome?	1	5	2.67 (1.33)
14. In the past 4 weeks, to what extent did mycosis fungoides or Sézary syndrome make you feel uncomfortable being around people other than close family and friends?	1	5	1.94 (1.20)

Identifying the Optimal Item Response Theory Measurement Model

Determining the optimal IRT model to calibrate the items was an iterative process based on empirical evidence and substantive rationale [11,12]. Since items were grouped into 2 rating scales, frequency and intensity, the Andrich-Grouped Rating Scale Model (G-RSM [12]) and Rating Scale Model (RSM [13]) were considered. Of note, the Partial Credit Model and Generalized Partial Credit Model were not considered as these models would likely produce unstable estimates due to the number of parameters to be estimated relative to the sample size.

To determine whether the rating scales for the intensity items and frequency items could be grouped, respectively, a partial credit model was used, and item characteristic curves (ICCs) were generated. These ICCs were similar within the groups (frequency and intensity). Next, a global chi-square test was performed to test whether the G-RSM significantly improved the fit above and beyond the RSM. Results revealed that the

G-RSM significantly improved model fit over the RSM (χ^2_3=8.7; *P*=.03).

Item Fit

Item fit was evaluated by examining item mean square infit and outfit statistics. There were 2 items that evidenced infit and outfit statistics above the commonly accepted cut-off of 1.33 [14] and were iteratively removed. Of note, these items still provide useful information about the patient experience and can be used in conjunction with the MF/SS-CTCL QoL global rating (see Multimedia Appendix 1). The remaining items evidenced adequate fit statistics and were retained for further analyses.

Unidimensionality

The assumption of unidimensionality was evaluated via an unrotated principal components analysis on the probability scale residuals in Winsteps [15]. Although the eigenvalue suggested the possible presence of a second dimension (eigenvalue=2.2), evaluation of item content and amount of variance explained by the Rasch measurement model (62.5%) provided support that the 12 items were measuring a unidimensional construct.

Evaluation of Rating Scale Performance

Andrich thresholds were examined to further ensure that the rating scales for the set of intensity items and the set of frequency items were performing as expected. Thresholds were ordered, indicating that a higher interference in QoL is required to endorse a higher frequency or intensity response category (Figures 1 and 2). Figure 1 depicts the relationship between interference with QoL and response option selection for the frequency items, whereby the different color curves represent the probability of selecting one of the response options. Specifically, “never,” “rarely,” “sometimes,” “often,” and “always” are represented by the red, blue, purple, gray, and green curves, respectively. This figure shows that a higher interference in the level of QoL is required to endorse a higher frequency.

Figure 2 depicts the relationship between interference with QoL and response option selection for the intensity items, whereby the different color curves represent the probability of selecting one of the response options. Specifically, “not at all,” “a little bit,” “somewhat,” “quite a bit,” and “very much” are represented by the red, blue, purple, gray, and green curves, respectively. This figure shows that a higher interference in the level of QoL is required to endorse a higher intensity.

Figure 1. Category response curves for the frequency items.

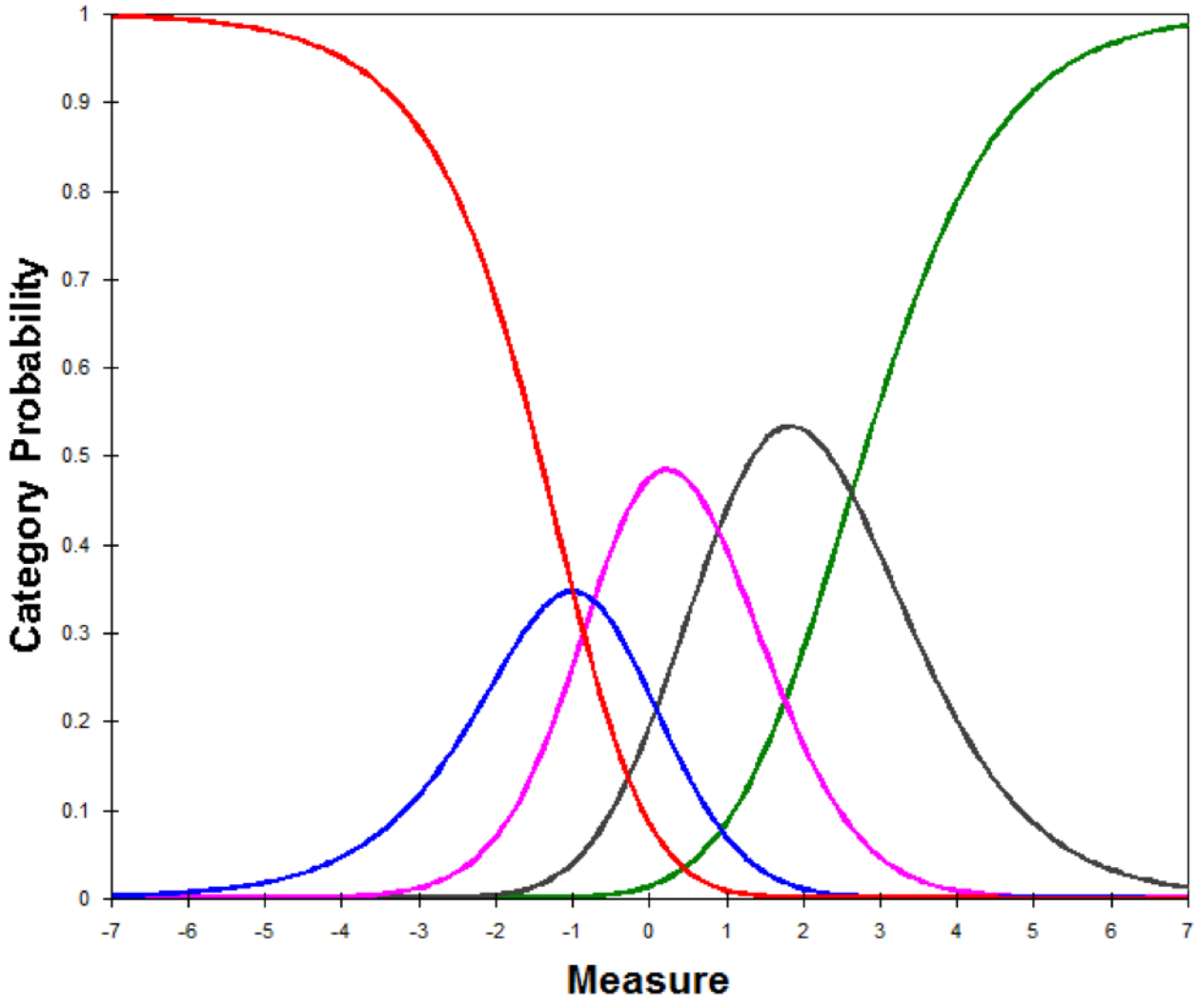
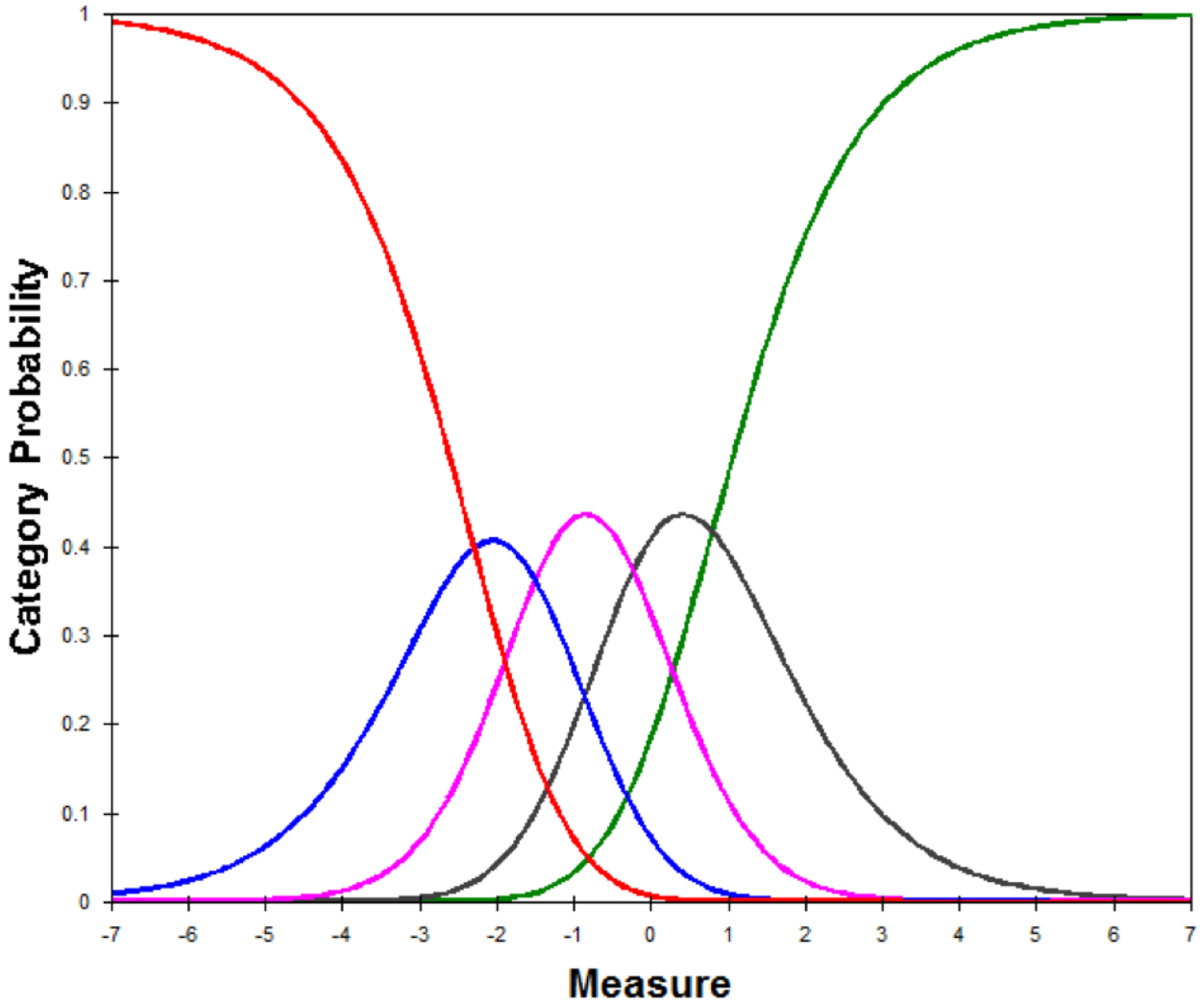


Figure 2. Category response curves for the intensity items.



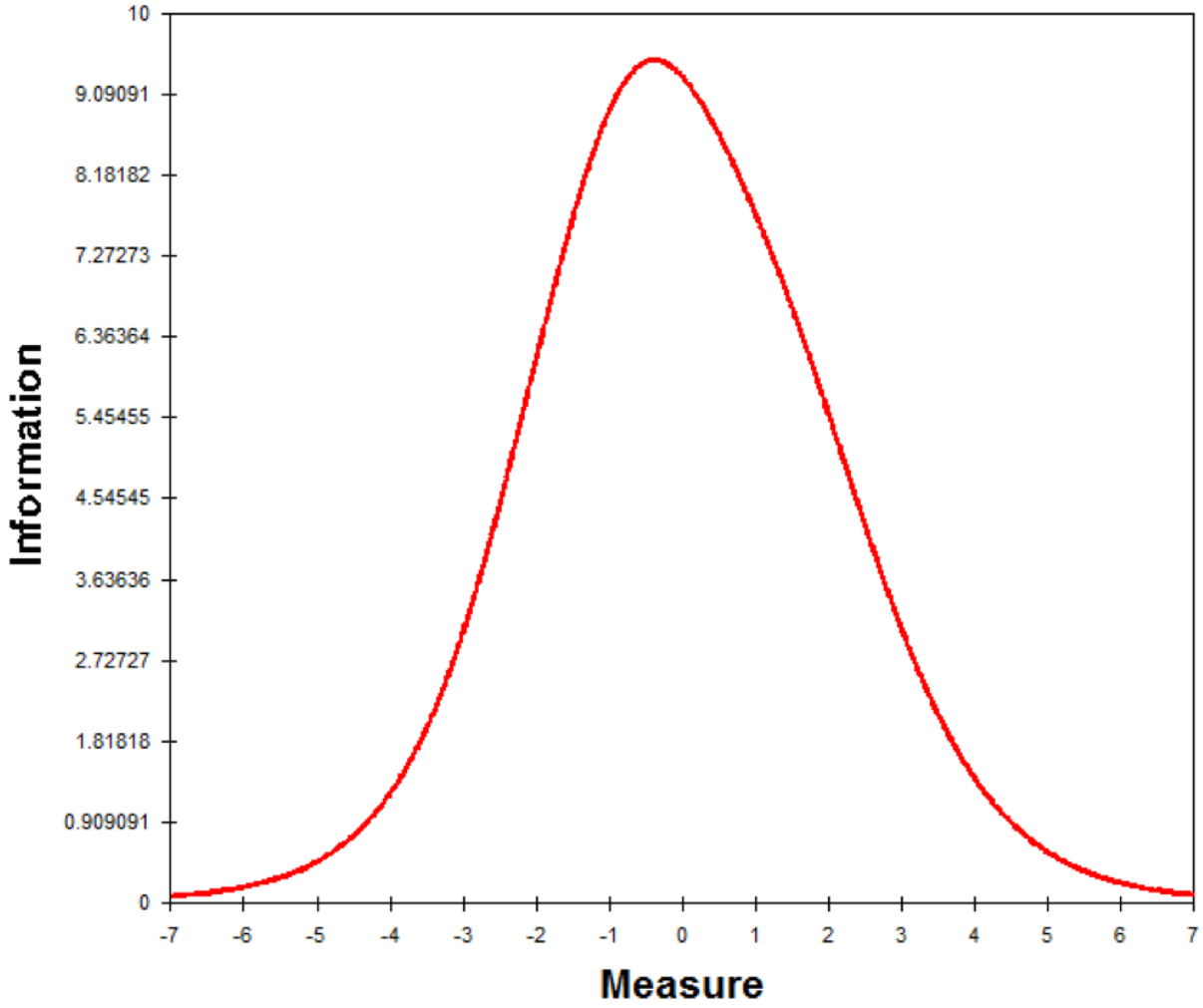
Reliability

Person reliability for the 12-item scale was 0.87, suggesting that the MF/SS-CTCL QoL is able to discriminate between individuals with low and high levels of interference in their QoL. Item reliability was 0.97, which suggests that the sample was large enough to locate items on QoL. Test-retest reliability (r=0.89; P<.001), calculated through a Pearson correlation between MF/SS-CTCL QoL scores at time 1 and time 2 (5 days later) revealed good to excellent stability.

Test Information

A test information curve was generated to evaluate the measurement precision of the MF/SS-CTCL QoL at various levels of the latent trait (QoL). The test information curve (Figure 3) provides evidence that the amount of interference with QoL was precisely estimated and that the MF/SS-CTCL QoL is best at differentiating people who have trait levels within about 2 SDs of the mean. Figure 3 depicts the amount of information (or precision of measurement) that is provided by the MF/SS-CTCL QoL measure across the latent construct of interference with QoL.

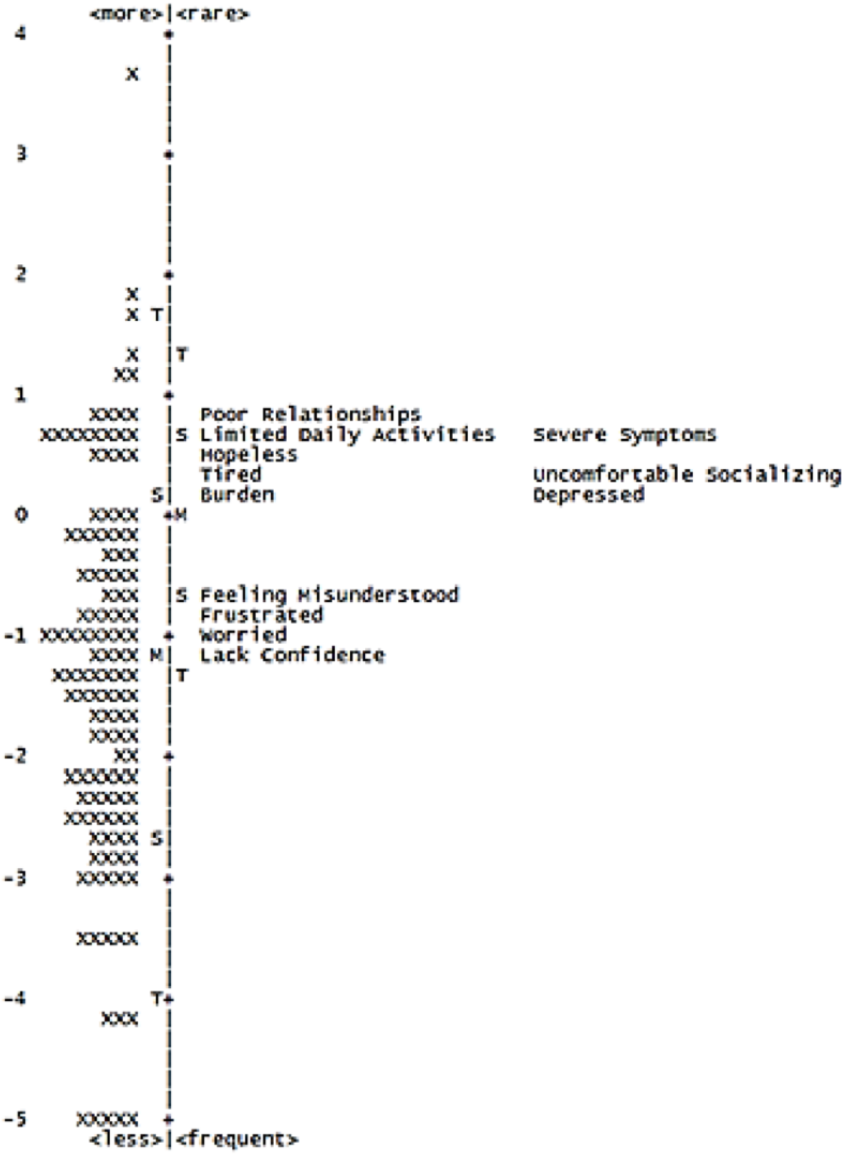
Figure 3. Test information curve.



Person-to-Item Map

Due to the unique properties of the Rasch model, it is possible to place both persons and items on the same interval-level scale or "ruler," depicted using a person-to-item map (Figure 4). This map can be interpreted as a vertical ruler, with persons (depicted on the left) and items (depicted on the right) ordered in relation to their difficulty or trait level using a scale (logits) with a mean of 0 and SD of 1. For example, on the MF/SS-CTCL QoL, "In the past 4 weeks, how confident did you feel about managing your mycosis fungoides or Sézary syndrome?" and "In the past 4 weeks, how much did you worry that your mycosis fungoides or Sézary syndrome may get worse?" were found to be easier (ie, require less impairment in QoL) to endorse. On the other hand, items near the top of the person-to-item map require a higher impairment in QoL to endorse (eg, "In the past 4 weeks, how much did mycosis fungoides or Sézary syndrome negatively affect your relationships with others close to you?" and "In the past 4 weeks, how much did your mycosis fungoides or Sézary syndrome limit your daily activities, ie, work inside and outside of the house, self-care such as cooking, cleaning, getting dressed, etc?"). This map is presented in Figure 4. Overall, examination of the map suggests adequate coverage of items across much of the latent trait. However, visual inspection of the map suggests limited person-to-item targeting at lower levels of interference with quality of life.

Figure 4. Person-to-item map.



Convergent and Discriminant Validity

To evaluate convergent and discriminant validity, a correlation matrix of the MF/SS-CTCL QoL, the Skindex-29, and syndrome stage was constructed. It was hypothesized that the MF/SS-CTCL QoL would be significantly more positively correlated with the Skindex-29, another QoL measure (convergent validity), than with syndrome stage (discriminant validity). The correlation between the MF/SS-CTCL QoL and the Skindex-29 (r=0.852; P<.001) was significantly greater than that between the MF/SS-CTCL QoL and syndrome stage (r=0.260; P<.001), providing support for convergent and discriminant validity.

Differential Item Function

DIF generally occurs when participants with an equal amount of the latent trait (interference in QoL) respond differently to an item [11]. DIF was assessed by gender, age, and race (white or nonwhite). DIF was considered notable if the DIF contrast estimate was >1.0 logit and significant at alpha=.05. Results revealed that the items did not have DIF at high enough levels to be considered problematic.

Scoring the Mycosis Fungoides/Sézary Syndrome Cutaneous T-Cell Lymphoma Quality of Life

A total raw MF/SS-CTCL QoL score is calculated by adding up the patient's total score from the 12 MF/SS-CTCL QoL items. Table 3 provides scaled scores (mean 100, SD 15) that correspond to the MF/SS-CTCL QoL total score. Although total raw scores of 10 or 11 are possible due to 2 items with the response choice "Does not apply (I don't have symptoms right now)," these scores should not be interpreted differently from a score of 12. For scoring purposes, "Does not apply (I don't have symptoms right now)" is scored as a 0. Furthermore, in order to score the MF/SS-CTCL QoL, each of the 12 items must be completed.

Table 3. Raw to scaled score conversion table.

Raw MF/SS-CTCL QoL[a] score	Scaled score[b]
≤12[c]	62
13	74
14	80
15	84
16	87
17	89
18	91
19	93
20	94
21	96
22	97
23	98
24	100
25	101
26	102
27	103
28	104
29	105
30	106
31	107
32	108
33	109
34	110
35	111
36	112
37	113
38	114
39	115
40	116
41	117
42	118
43	119
44	120
45	121
46	123
47	124
48	125
49	126
50	128
51	129
52	131

Raw MF/SS-CTCL QoL[a] score	Scaled score[b]
53	133
54	135
55	137
56	139
57	143
58	147
59-60	154

[a]MF/SS-CTCL: mycosis fungoides/Sézary syndrome cutaneous T-cell lymphoma quality of life.

[b]Scaled scores were standardized on the current sample to have a mean of 100 and an SD of 15.

[c]While it is possible to obtain a raw score of 10 or 11 due to endorsing "Does not apply (I don't have symptoms right now)" to MF/SS-CTCL QoL items, these scores should be viewed as equivalent to a 12.

Table 4. Qualitative descriptions of mycosis fungoides/Sézary syndrome cutaneous T-cell lymphoma quality of life scaled scores.

Scaled score	Description
62 to 89	No to low interference
91 to 105	Mild interference
106 to 117	Moderate interference
118 to 133	Substantial interference
135 to 154	Severe interference

Mycosis Fungoides/Sézary Syndrome Cutaneous T-Cell Lymphoma Quality of Life Interpretation

The qualitative description of scaled scores, provided in Table 4, was based on evaluating the distribution of scaled scores relative to the response categories. For example, individuals with scaled scores that corresponded with an average rating of 1 (not at all or never) across the items were described as having no to low interference, individuals with scaled scores that corresponded with an average rating of 2 (a little or rarely) were described as having mild interference, and individuals with scaled scores that corresponded with an average rating of 3 (somewhat or sometimes) across the items were described as having moderate interference (Table 4).

Discussion

Principal Findings

This research utilized a multistage instrument development process that incorporated both qualitative and quantitative components, including (1) development of a conceptual model through literature review and input from KOLs; (2) refinement of the conceptual model and generation of preliminary items through concept elicitation with patients; (3) item revisions based on feedback from patients during cognitive debriefing; and (4) empirical testing to evaluate psychometric functioning and finalize the MS/SS-CTCL QoL. The results provide strong support for reliability and validity of the MS/SS-CTCL QoL. Specifically, results indicate that the rating scale was functioning as expected, and the 12-item MS/SS-CTCL QoL exhibited adequate person reliability, excellent test-retest reliability, high levels of measurement precision, good person-to-item targeting, and evidence of convergent and discriminant validity. Items did not evidence significant bias based on gender, age, or race.

This study used state-of-the-art modern test theory approaches, which are considered the "gold standard" in test construction methodology as they rely on stronger measurement assumptions and produce more reliable results than classical approaches [11,16]. Further, Rasch modeling allows for new items to be incorporated into the instrument without having to establish the validity of the entire bank. This advantage may be particularly important as treatments improve and disease management changes over time.

Limitations

The sample of patients with MF who participated in this validation study was largely of those with stage I disease. Incorporating a greater number of patients who represent the more advanced stages in the item generation process may have resulted in different item content. Consequently, gathering feedback from patients with more advanced stages will likely be a critical part of future instrument refinement.

All data collected from patients during this study relied exclusively on patient report, and patient diagnosis and stage could not be verified by a licensed medical professional. Additionally, patients were recruited from the internet, potentially excluding patients who do not have internet access or those whose health or functioning may interfere with their ability to use the internet. Similarly, many of the participants for this research were recruited through PatientsLikeMe, potentially limiting the generalizability of findings. For example, members of the PatientsLikeMe Web-based community may be more conscious, engaged in their health, and comfortable

sharing health information than the general population [17]. Therefore, future research may evaluate whether the items function differently among patients recruited through different sources, such as hospitals or university clinics.

Despite the partnership from the Cutaneous Lymphoma Foundation to assist with recruitment of patients with this rare disease, obtaining sample sizes adequate for each phase of measure development was challenging, and the same patients participated in several stages of the development process. Additional studies should be performed to replicate these findings. Future research might also evaluate this instrument's ability to detect change over time as a patient's stage, treatment, or health status changes.

Finally, evaluation of person-to-item targeting suggested that this scale may be limited in its ability to differentiate persons who may be experiencing little to no interference with QoL. However, from a clinical perspective, this is likely not problematic. That is, clinicians may be less concerned with precisely measuring interference with QoL and differentiating between patients who are very low on interference with QoL than with precisely measuring and tracking interference with QoL in patients who are experiencing some level of interference.

Conclusions

The MF/SS-CTCL QoL is the first MF/SS-specific instrument to capture the impact of MF/SS-CTCL on patients' health-related QoL. Incorporating the patient perspective throughout the development process likely increased the relevancy of MF/SS-CTCL QoL content for this patient population. The MF/SS-CTCL QoL was developed in partnership with the Cutaneous Lymphoma Foundation with the intention of improving care for MF/SS-CTCL patients. Therefore, the MF/SS-CTCL QoL is free for clinicians, patients, and researchers and can be downloaded free of charge from the ORE.

Acknowledgments

The authors would like to thank Jean-Philip Okhovat and Marcy Fitz-Randolph for their contributions to the manuscript and Magdalena Hurrington, Kristina Simacek, and Auriell Towner for their contributions to this research.

Authors' Contributions

JB, MN, MS, and GS designed the study and provided input throughout the study. JB collected the data. YHK provided clinical expertise throughout the study and assisted with the finalization of the instrument. SM, RAB, and JB analyzed portions of the data. SM wrote the manuscript along with contributions from all the authors. All the authors read and approved the final manuscript.

Conflicts of Interest

This research was funded by Actelion US, Inc. Authors MN, MS, and GS were employees of Actelion US, Inc at the time this research was conducted.

Multimedia Appendix 1

The MF/SS-CTCL QoL instrument.

[PDF File (Adobe PDF File), 43KB - jmir_v21i1e11302_app1.pdf]

References

1. Jawed SI, Myskowski PL, Horwitz S, Moskowitz A, Querfeld C. Primary cutaneous T-cell lymphoma (mycosis fungoides and Sézary syndrome): part I. Diagnosis: clinical and histopathologic features and new molecular and biologic markers. J Am Acad Dermatol 2014 Feb;70(2):205.e1-205.e16. [doi: 10.1016/j.jaad.2013.07.049] [Medline: 24438969]
2. Olsen E, Vonderheid E, Pimpinelli N, Willemze R, Kim Y, Knobler R, ISCL/EORTC. Revisions to the staging and classification of mycosis fungoides and Sezary syndrome: a proposal of the International Society for Cutaneous Lymphomas (ISCL) and the cutaneous lymphoma task force of the European Organization of Research and Treatment of Cancer (EORTC). Blood 2007 Sep 15;110(6):1713-1722 [FREE Full text] [doi: 10.1182/blood-2007-03-055749] [Medline: 17540844]
3. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [Medline: 8433390]
4. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol 1993 Mar;11(3):570-579. [doi: 10.1200/JCO.1993.11.3.570] [Medline: 8445433]
5. Desai N, Poindexter G, Monthrope Y, Bendeck S, Swerlick R, Chen S. A pilot quality-of-life instrument for pruritus. J Am Acad Dermatol 2008 Aug;59(2):234-244. [doi: 10.1016/j.jaad.2008.04.006] [Medline: 18550210]
6. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. Clin Exp Dermatol 1994 May;19(3):210-216. [Medline: 8033378]

7. Chren MM, Lasek RJ, Quinn LM, Mostow EN, Zyzanski SJ. Skindex, a quality-of-life measure for patients with skin disease: reliability, validity, and responsiveness. J Invest Dermatol 1996 Nov;107(5):707-713 [FREE Full text] [Medline: 8875954]
8. Reich A, Heisig M, Phan NQ, Taneda K, Takamori K, Takeuchi S, et al. Visual analogue scale: evaluation of the instrument for the assessment of pruritus. Acta Derm Venereol 2012 Sep;92(5):497-501 [FREE Full text] [doi: 10.2340/00015555-1265] [Medline: 22102095]
9. Demierre MF, Gan S, Jones J, Miller DR. Significant impact of cutaneous T-cell lymphoma on patients' quality of life: results of a 2005 National Cutaneous Lymphoma Foundation Survey. Cancer 2006 Nov 15;107(10):2504-2511 [FREE Full text] [doi: 10.1002/cncr.22252] [Medline: 17048251]
10. Kerr C, Nixon A, Wild D. Assessing and demonstrating data saturation in qualitative inquiry supporting patient-reported outcomes research. Expert Rev Pharmacoecon Outcomes Res 2010 Jun;10(3):269-281. [doi: 10.1586/erp.10.30] [Medline: 20545592]
11. Embretson SE, Reise SP. Item response theory for psychologists. In: Mahwah, NJrlbaum Publishers. Mahwah, NJ: Erlbaum Publishers; 2000.
12. Reise S, Revicki D. eds. Handbook of item response theory modeling: Applications to typical performance assessment. London, United Kingdom: Routledge; 2014.
13. Andrich D. Application of a Psychometric Rating Model to Ordered Categories Which Are Scored with Successive Integers. Applied Psychological Measurement 1978 Oct 01;2(4):581-594. [doi: 10.1177/014662167800200413]
14. Wilson M. Constructing measures: An item response modeling approach. 4th ed. New York, NY: Lawrence Erlbaum Associates; 2004.
15. Bond T, Fox C. Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers; 2015.
16. An X, Yung Y. SAS Institute, Inc. 2014. Item Response Theory: What It Is and How You Can Use the IRT Procedure to Apply It URL: https://support.sas.com/resources/papers/proceedings14/SAS364-2014.pdf [accessed 2016-01-06]
17. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]

Abbreviations

CTCL: cutaneous T-cell lymphoma
DIF: differential item function
G-RSM: Grouped Rating Scale Model
ICC: item characteristic curve
IRT: item response theory
KOL: key opinion leader
MF: mycosis fungoides
ORE: Open Research Exchange
PRO: patient-reported outcome
QoL: quality of life
RSM: Rating Scale Model
SS: Sézary syndrome

Edited by G Eysenbach; submitted 20.06.18; peer-reviewed by O Segal, M Schoen; comments to author 30.08.18; revised version received 23.09.18; accepted 23.09.18; published 07.01.19.

Please cite as:
McCaffrey S, Black RA, Nagao M, Sepassi M, Sharma G, Thornton S, Kim YH, Braverman J
Measurement of Quality of Life in Patients with Mycosis Fungoides/Sézary Syndrome Cutaneous T-Cell Lymphoma: Development of an Electronic Instrument
J Med Internet Res 2019;21(1):e11302
URL: https://www.jmir.org/2019/1/e11302/
doi:10.2196/11302
PMID:30617041

Original Paper

An Electronic Patient-Reported Outcome Tool for the FACT-B (Functional Assessment of Cancer Therapy-Breast) Questionnaire for Measuring the Health-Related Quality of Life in Patients With Breast Cancer: Reliability Study

Lina Maria Matthies[1*]; Florin-Andrei Taran[2*]; Lucia Keilmann[1]; Andreas Schneeweiss[1]; Elisabeth Simoes[3]; Andreas D Hartkopf[2]; Alexander N Sokolov[3,4]; Christina B Walter[2]; Nina Sickenberger[1]; Stephanie Wallwiener[1]; Manuel Feisst[5]; Paul Gass[6]; Michael P Lux[6]; Florian Schuetz[1]; Peter A Fasching[6]; Christof Sohn[1]; Sara Y Brucker[3]; Joachim Graf[3,4*]; Markus Wallwiener[1*]

[1]Hospital for General Obstetrics and Gynecology, Gynecologic Oncology, National Center for Tumor Diseases, University Hospital Heidelberg, Heidelberg, Germany

[2]Department of Women's Health, University Hospital Tübingen, Tübingen, Germany

[3]Research Institute for Women's Health, Department of Women's Health, University Hospital Tübingen, Tübingen, Germany

[4]Section of Midwifery Science, Institute for Health Sciences, University Hospital Tübingen, Tübingen, Germany

[5]Institute for Medical Biometry and Informatics, University Hospital Heidelberg, Heidelberg, Germany

[6]Department of Gynecology and Obstetrics, University Breast Center Franconia, Comprehensive Cancer Center Erlangen-EMN, University Hospital Erlangen, Erlangen, Germany

[*]these authors contributed equally

Corresponding Author:
Markus Wallwiener
Hospital for General Obstetrics and Gynecology
Gynecologic Oncology, National Center for Tumor Diseases
University Hospital Heidelberg
Im Neuenheimer Feld 440
Heidelberg, 69120
Germany
Phone: 49 (0)6221 5636956
Email: markus.wallwiener@gmail.com

Abstract

Background: The most frequent malignant disease in women is breast cancer. In the metastatic setting, quality of life is the primary therapeutic goal, and systematic treatment has only a limited effect on survival rates; therefore, the concept of the health-related quality of life (HRQoL) and measurement of patient-reported outcomes (PROs) are gaining more and more importance in the therapy setting of diseases such as breast cancer. One of the frequently used questionnaires for measuring the HRQoL in patients with breast cancer is the Functional Assessment of Cancer Therapy-Breast (FACT-B). Currently, paper-based surveys still predominate, as only a few reliable and validated electronic-based questionnaires are available. ePRO tools for the FACT-B questionnaire with proven reliability are missing so far.

Objective: The aim of this study was to analyze the reliability of tablet-based measurement of FACT-B in the German language in adjuvant (curative) and metastatic breast cancer patients.

Methods: Paper- and tablet-based questionnaires were completed by a total of 106 female adjuvant and metastatic breast cancer patients. All patients were required to complete the electronically based (ePRO) and paper-based version of the FACT-B. A frequency analysis was performed to determine descriptive sociodemographic characteristics. Both dimensions of reliability (parallel forms reliability using Wilcoxon test and test of internal consistency using Spearman ρ) and agreement rates for single items, Kendall tau for each subscale, and total score were analyzed.

Results: High correlations were shown for both dimensions of reliability (parallel forms reliability and internal consistency) in the patients' response behavior between paper-based and electronically based questionnaires. Regarding the reliability test of

parallel forms, no significant differences were found in 35 of 37 single items, while significant correlations in the test for consistency were found in all 37 single items, in all 5 sum individual item subscale scores, as well as in total FACT-B score.

Conclusions: The ePRO version of the FACT-B questionnaire is reliable for patients with breast cancer in both adjuvant and metastatic settings, showing highly significant correlations with the paper-based version in almost all questions all subscales and the total score.

(J Med Internet Res 2019;21(1):e10004) doi:10.2196/10004

KEYWORDS

breast cancer; ePRO measurement; FACT-B; HRQoL; patient-reported outcomes; reliability of ePRO

Introduction

Breast Cancer: Epidemiological Relevance

The most frequent malignant disease in women is breast cancer; indeed, about 70,000 new cases of breast cancer are diagnosed in Germany every year. Therapeutic options have been improved, resulting in an overall 5-year survival rate of patients with early-stage disease of >90% [1-3]. In contrast, the prognosis of metastatic breast cancer is significantly poorer, since palliative treatment often remains the only option due to the low probability of cure in patients with metastatic disease [4,5].

Patient-Reported Outcomes and Health-Related Quality of Life in Patients With Breast Cancer

Since systematic palliative treatment has only a limited effect on survival rates, the concept of the health-related quality of life (HRQoL) and measurement of patient-reported outcomes (PROs) are gaining more and more importance in the therapy setting of progressive diseases, such as breast cancer, especially in the adjuvant or metastatic setting [6-10]. Patients with cancer often suffer from symptoms and adverse events during their treatment, which are sometimes underestimated in clinical routine [11-13]. Although clinicians' perception of symptoms can predict unfavorable clinical events more precisely, patients´ reports can reflect the daily health status more adequately [14]. PROs comprise various aspects of the subjectively perceived state of health from the patients' point of view such as HRQoL, satisfaction with care, and drug adherence [9,15,16]. With regard to the therapy setting, monitoring HRQoL and the occurrence of symptoms appears to be of particular relevance, primarily during therapy, but also as a long-term follow-up for improving and supporting patients´ well-being [17-20]. The importance of measuring PRO in patients with breast cancer is also stated in the German S3-guideline [21]. As (metastatic) breast cancer often remains an incurable disease with only palliative treatment options, monitoring the HRQoL is highly relevant in these patients [4,22,23]. Different questionnaires highlight different aspects of symptoms and HRQoL [24]. One of the frequently used questionnaires for measuring the HRQoL in patients with breast cancer is the Functional Assessment of Cancer Therapy-Breast (FACT-B), a validated, multidimensional questionnaire with 37 items that build 5 dimensions (subscales) when using a 5-point Likert scale [25-27].

Electronic Measurement of Patient-Reported Outcomes

Collecting and analyzing pencil and paper-based data are difficult tasks without possibilities for direct response or interaction [15]; therefore, the electronic-based measurement of PRO (ePRO) is gaining more and more importance. There are several advantages of ePRO such as rapid access to data, a probable avoidance of errors during data entry, fewer missing data in comparison with paper-based surveys, the capacity to trigger alerts or notifications for answers to special circumstances, and an improvement in patients' willingness to report sensitive information [23,28]. Although paper-based surveys of PRO still predominate because there are only a few reliable and validated ePRO questionnaires, numerous projects have evaluated the feasibility and acceptance of HRQoL in the ePRO measurement in the last few years [29-34]. Nevertheless, knowledge regarding patients' acceptance, feasibility, and barriers remains limited [35], especially because hurdles might exist in relation to health status, technical skills, and socioeconomic aspects, which could influence both patients´ willingness to use ePRO and their response behavior [10,36,37]. Although studies have already demonstrated a potential equivalence between some paper-based PRO and ePRO, the reliability of ePRO questionnaires should be verified so as not to endanger the validity of ePRO surveys [10,36-40]. Indeed, ePRO tools for the FACT-B questionnaire with proven reliability are missing so far.

Aims and Objectives

The aim of the study was to analyze the reliability of a tablet-based ePRO app for FACT-B in German for measuring the HRQoL in adjuvant and metastatic breast cancer patients in comparison with the validated paper-based version of FACT-B. It was planned to determine whether differences exist in response behavior between the validated paper-based PRO version of FACT-B and a new ePRO version, whether the answers between paper-based and ePRO questionnaire differ in a relevant way, and whether the patients' response behavior is influenced by the mode of answering (paper- or tablet-based). In order to achieve these aims, patients were asked to complete both the paper- and tablet-based version of the FACT-B questionnaire.

Methods

Sample and Study Design

From July 2015 to May 2016, paper-based and tablet-based PRO questionnaires were completed by a total of 106 female adjuvant and metastatic breast cancer patients treated consecutively at the Department of Women's Health (Tübingen, Germany) and the National Cancer Center (Heidelberg, Germany). Patients were recruited as a part of the electronic

Patient-Reported Outcomes and Compliance Analysis (ePROCOM) and Patient Engagement Pilotstudie Mammakarzinom-individualisierte und Ressourcen-effiziente Patient Reported Outcomes Erfassung durch Digitale Therapieunterstützungssysteme (PEPPER) studies. The aims of ePROCOM were to evaluate the general patient acceptance and practicability of a Web-based app for a PRO questionnaire for patients with adjuvant or metastatic breast cancer. Patients were asked to participate to compare their response behavior in paper-based and Web-based questionnaires and analyze the reliability of the ePRO versions of the questionnaires European Organisation for Research and Treatment of Cancer (EORTC) quality of life questionnaire (QLQ)-C30, as published previously [40], and FACT-B (reported in this paper). The PEPPER study intends to evaluate the impact of Web-based and paper-based PRO for health care services by patients from the Prospective Academic Translational Research Network for the Optimization of the Oncological Health Care Quality in the Adjuvant and Advanced/Metastatic Setting: Health Care Research, Pharmacogenomics, Biomarkers, Health Economics study. The methods are described in detail in the EORTC paper [40]. The inclusion criteria of the ePROCOM and PEPPER studies were female gender, full legal age, the proven diagnosis of breast cancer in adjuvant or metastatic setting, sufficient language skills in German, and signed declaration of consent. The exclusion criterion was participation in other studies to minimize the burden of questionnaires. Patients were asked to complete the questionnaire during an outpatient visit at the hospital under the supervision of an attending physician. The study was designed as a 2-center (Tübingen and Heidelberg), 2-armed, prospective randomized trial. All patients were required to complete both the electronically based (ePRO) and paper-based version of the FACT-B HRQoL questionnaire. Patients in arm A were assigned to start with the tablet version followed by the paper version in the same session. Patients in arm B filled out the paper-based version followed by the tablet-based questionnaire. The randomization procedure was based on permuted-block randomization, which strives to generate equally large treatment groups [41,42]. The postexposure acceptance for using the ePRO tool was high (92%) as patients were asked whether they could potentially imagine using tablet-based tools before using ePRO [37]. Patients were informed about the aims of this study and were asked for their consent *ex ante*. The study was approved by the Ethics Committee at the University of Tübingen (project number 089/2015B02) [40].

Procedure

The data collection was performed in 4 parts. The first part focused on patients socioeconomic variables. The second part comprised the FACT-B questionnaire, consisting of 37 questions with responses required on a 5-point Likert scale (from 0, *not at all* through 4, *very much*) that constitute 5 dimensions [25-27,43]. The response options labels (*not at all* through *very much*; Figure 1) were the same in both the standard German paper-based and the tablet-based versions. While in the third part of the assessment, patients were asked about preexisting technical skills, their willingness to use ePRO, and potential barriers in relation to their health status [37], the fourth part concerned with the patients' evaluation of the ePRO tool (manuscript in preparation). All patients completed the second part of the assessment both on paper and using a tablet, while they answered the questions in the other parts only in paper-based form. In this study, we report the results of the second part of the assessment (reliability analysis of the ePRO tool of FACT-B) [40].

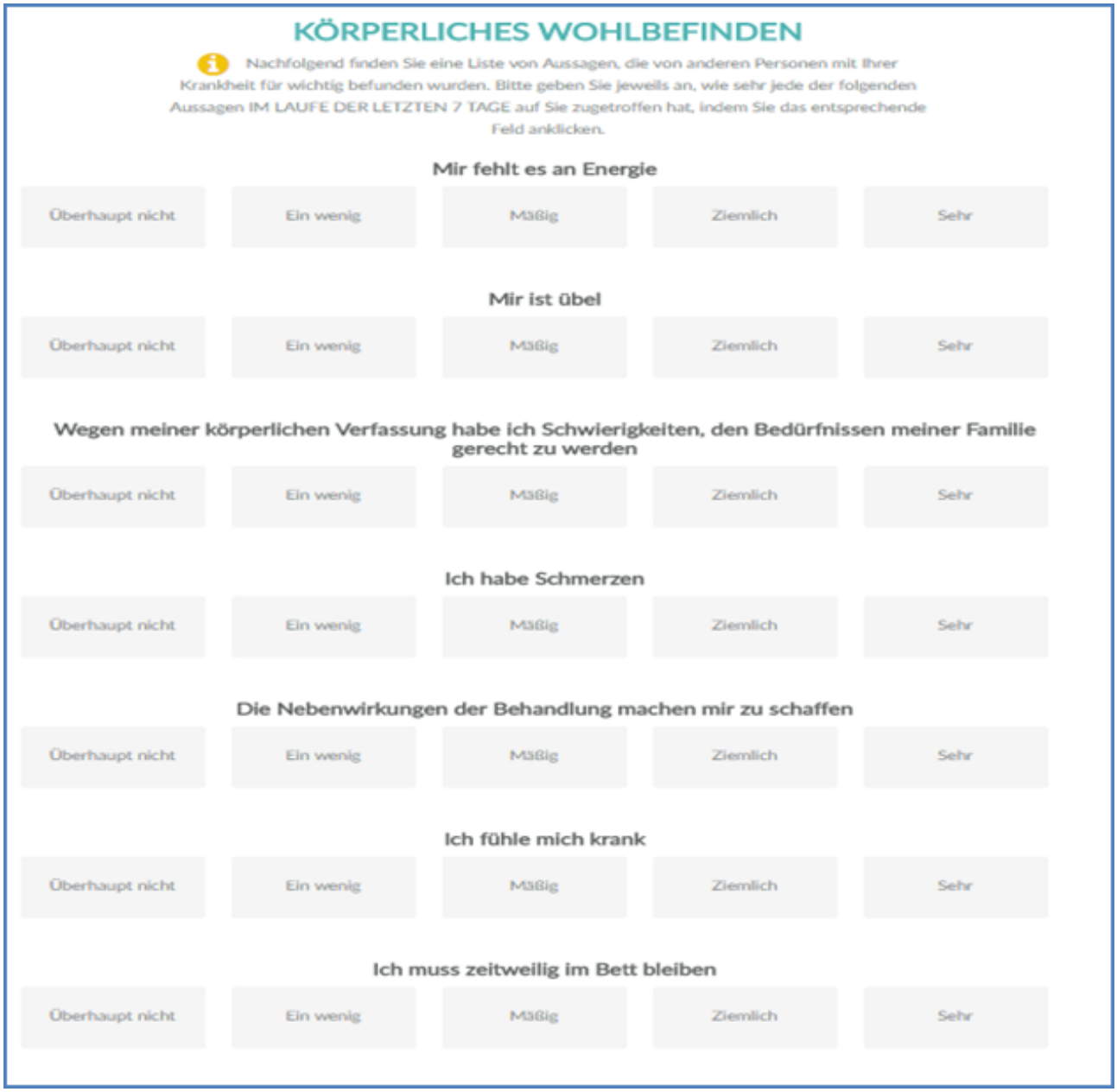
Specifics of the Electronic Patient-Reported Outcomes Tool

For measuring ePRO, we used the "Patient-informiert-interaktiv-Arzt (PiiA)," that is "patient interactively informs doctor" Web-based app, which allows patients to answer the relevant questions on a tablet. The PiiA portal is a Web-based solution for capturing PROs, which our working group has developed. Patients receive anonymized user credentials and are asked to complete the FACT-B questionnaire. Figure 1 shows the user interface of the first set of questions of the German FACT-B. After completing the questionnaires, patients log out and their pseudoanonymized data are backed up on a local storage device and securely locked [40].

Statistical Analyses

All statistical analyses were conducted using SPSS Statistics (IBM, version 24). First, a frequency analysis was performed to determine the descriptive sociodemographic characteristics of patients. After that, we analyzed both dimensions of reliability (parallel forms reliability and test of internal consistency) and examined the disparity of responses and the rate of consistency between paper- and tablet-based responses. Both types of reliability were calculated for the 37 single items as well as for scores of the 5 dimensions, including the subscales for Physical Well-Being (PWB), Social/Family Well-Being (SWB), Emotional Well-Being (EWB), Functional Well-Being (FWB), and Breast Cancer Subscale (BCS), and the FACT-B total score in accordance with the FACT-B guidelines [25-27]. According to the Shapiro-Wilks test, the paired samples were not normally distributed; therefore, we used the Wilcoxon test to identify possible statistically significant differences in the *test of parallel forms reliability* both between the single items and the scores. Initially, the mean values of the paper- and tablet-based measures were calculated according to the official FACT-B guidelines [25-27]. Second, the consistency analyses were performed by calculation of Spearman rank correlation coefficient (Spearman ρ) and agreement rates for every FACT-B item along with rank correlation (Kendall tau) for each scale. Prior to that, we performed chi-square tests and Shapiro-Wilks test comparing metastatic and adjuvant breast cancer patients to identify possible statistically significant differences in relation to HRQoL. In all analyses, $P<.05$ (2-tailed) was considered indicative of statistically significant differences (alpha=.05).

Figure 1. Screenshot of the Patient-informiert-interaktiv-Arzt app's FACT-B (Functional Assessment of Cancer Therapy-Breast) questionnaire for the dimension "physical well-being" (German). Source: Authors work, licensed under fair use.



As such an analysis is considered an explorative study, all reported *P* values can be taken as purely descriptive. Missing values (which arose when patients did not answer individual questions) were appropriately taken into account in the calculation of the scores that were ignored in the statistical calculation [40]. Both figures (boxplot and correlation diagram) were generated in SPSS 24.

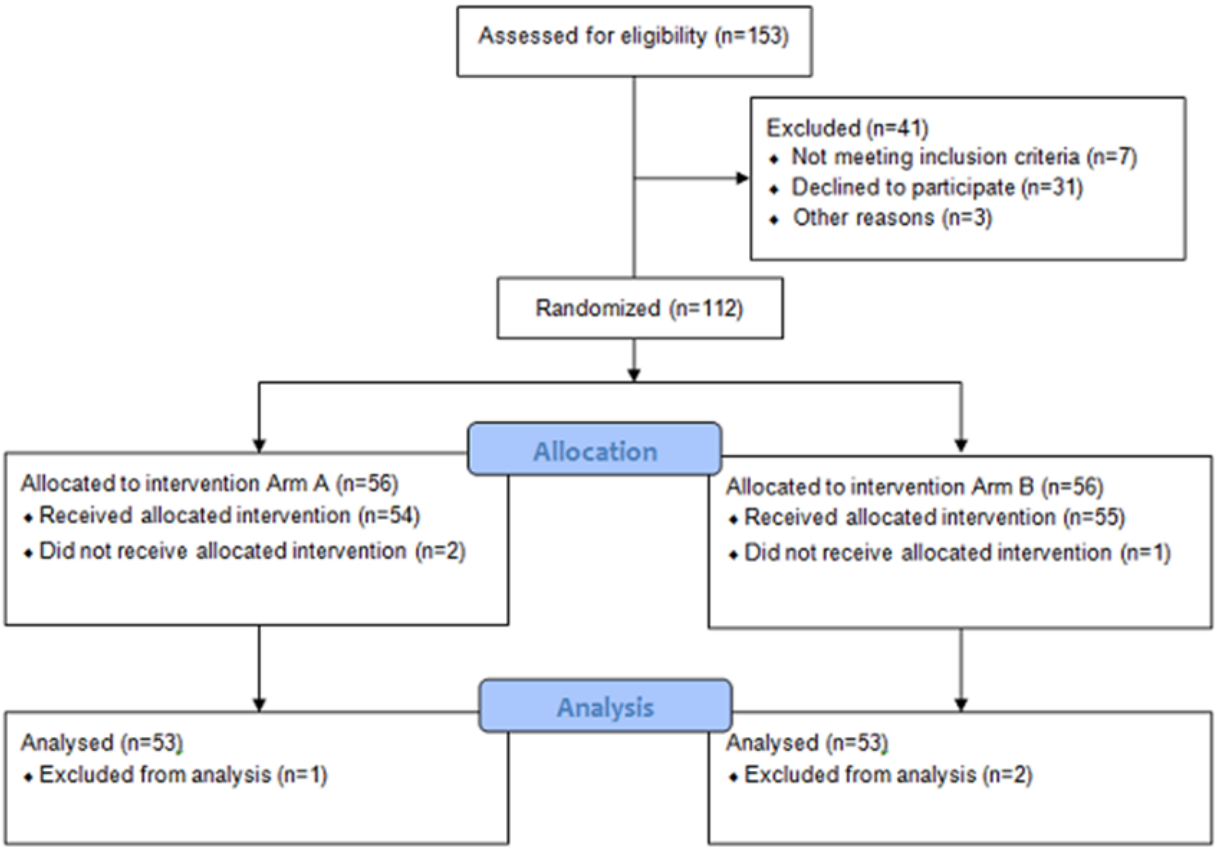
Results

Patient Enrollment

Overall, 106 eligible female patients with breast cancer were recruited who completed questions from FACT-B both in a paper-based format and electronically using a tablet. Originally, 153 patients were assessed for eligibility, of which 47 were excluded during recruiting, allocation, and data analysis, as shown in the Consolidated Standards of Reporting Trials flow diagram (Figure 2).

In all, 53 patients were assigned to tablet-based filling followed by paper-based filling in the same session (arm A), while the same number of patients completed the paper-based version followed by the tablet-based questionnaire (arm B). Both the paper- and tablet-based questionnaires were completed by patients consecutively during the same ambulance visit. Patients who had not completed more than half of the FACT-B questions in either paper- or tablet-based format were excluded (arm A, 1 patient; arm B, 2 patients). We did not find any significant differences between the 2 arms in the response behavior, sociodemographic status, or therapy setting; therefore, the 2 arms were considered together. Beforehand, all single items of the 2 arms were compared. Ten patients (arm A) and 16 patients (arm B), respectively, produced missing data in some questions (more often in the tablet-based questionnaire) [40].

Figure 2. Consolidated Standards of Reporting Trials flow diagram.



Sociodemographic Variables

Tables 1 and 2 shows the sociodemographic characteristics of the study group, with 71.6 (76/106) of patients in adjuvant therapy and 28.3% (30/106) in a metastatic situation. We did not find any marked intergroup differences between metastatic and adjuvant breast cancer patients in either ePRO or paper-based PRO. Although adjuvant and metastatic patients differed in age and HRQoL, as metastatic patients were older and reported a poorer HRQoL, we found no differences in the ePRO response behavior between the 2 groups. In addition, there were no differences in the reliability in any of the single items and scales between metastatic and adjuvant patients; hence, the complete patient collective was considered together. The mean age of the whole collective was 51.0 years; nearly one-third of the patients had a higher level of education (high school diploma) [40].

Table 1. Sociodemographic characteristics of the patients (n=106).

Sociodemographic variables	Descriptive analyses
Age	
Mean (SD)	51 (11.31)
Median (range, minimum-maximum)	52 (54, 30-84)
Level of education[a]	
Median	3
Interquartile range (25%-75% quantiles)	2 (3-5)

[a]Level of education: 1=lowest; 5=highest.

Table 2. Education level and therapy setting of the patients (n=106).

Variable	Frequency, n (%)	95% CI
Level of education		
No qualification	1 (0.9)	0-6
Main or secondary school graduation	43 (40.5)	32-50
Advanced technical graduation	19 (17.9)	10-26
High school diploma[a]	33 (31.1)	22-40
Not specified	10 (9.4)	2-15
Therapy setting		
Metastatic	30 (28.3)	19-35
Adjuvant setting	76 (71.6)	61-83

[a]High school diploma indicates "Abitur."

Parallel Forms Reliability

Table 3 presents the results of the Wilcoxon test for analyzing parallel forms reliability in the single items of FACT-B. The ePRO tool seems to demonstrate acceptable parallel forms reliability as only 2 significant differences (out of 37 in total) could be found in the single-item comparison. A weak statistically significant difference could only be identified in questions GP 4 (*I have pain*) and GS 2 (*I get emotional support from my family*). The pain reports were slightly higher in the ePRO questionnaire (*P*=.012), while emotional support was evaluated somewhat higher in paper-based PRO (*P*=.036). The differences were only slight, though, as the medians of the 2 items could not be distinguished from each other. In 35 of 37 questions in the FACT-B, no statistically significant differences were observed between patients' answers in the paper-based questionnaire and ePRO.

In addition, slight differences were noted between paper-based PRO and ePRO in the individual item scores of the 5 dimensions and the total FACT-B score (Table 4). The differences were significant in 2 dimensions (EWB and BCS) but not in the total scores because the scores included several questions (PWB, SWB, and FWB: 7 questions; EWB: 6 questions; BCS: 10 questions), whereby differences in patients' responses to single questions could be multiplied in the scores and missing values caused fluctuations between the paper- and tablet-based PRO. The statistical differences in 2 scores, therefore, appear to be of methodological origin rather than attributable to differences in the response behavior. As only 5 response options (0-4) are available for each question in the FACT-B questionnaire, the ranking procedure often results in a large number of ties in the Wilcoxon test, which produces larger *P* values.

The total score is slightly higher in ePRO (mean difference: 1.73; median difference: 0.63), but without statistically significant differences. Figure 3 shows the distribution of the paper-based and ePRO total score for FACT-B in a boxplot. It is obvious that the whisker but not the interquartile range is broader in the paper-based version.

Table 3. Parallel forms reliability (Wilcoxon test) in single items.

Single items	Paper-based patient-reported outcomes		Electronic patient-reported outcomes		*P* value
	Mean (SD)	Median (Interquartile range)	Mean (SD)	Median (Interquartile range)	
Physical Well-Being					
GP1	1.68 (1.22)	2 (2)	1.59 (1.21)	1 (1)	.22
GP2	0.54 (0.86)	0 (1)	0.60 (0.88)	0 (1)	.26
GP3	1.34 (1.21)	1 (2)	1.29 (1.18)	1 (2)	.72
GP4	0.98 (1.01)	1 (1.25)	1.03 (0.99)	1 (2)	*.01* [a]
GP5	1.50 (1.17)	1 (1)	1.58 (1.09)	1 (1)	.08
GP6	1.22 (1.12)	1 (2)	1.23 (1.05)	1 (2)	.24
GP7	0.77 (1.09)	0 (1)	0.73 (1.03)	0 (1)	>.99
Social/Family Well-Being					
GS1	3.20 (0.94)	3 (1)	3.16 (1.04)	3 (1)	.82
GS2	3.61 (0.73)	4 (1)	3.5 (0.86)	4 (1)	*.04* [a]
GS3	3.18 (1.09)	4 (1)	3.14 (1.1)	3 (1)	>.99
GS4	3.33 (0.75)	3 (1)	3.28 (0.81)	3 (1)	.39
GS5	3.37 (0.82)	4 (1)	3.33 (0.77)	3 (1)	.11
GS6	3.65 (0.77)	4 (.0)	3.63 (0.82)	4 (1)	.83
GS7	1.98 (1.66)	2 (2.5)	2.07 (1.11)	2 (2)	.72
Emotional Well-Being					
GE1	1.32 (1.09)	1 (1)	1.22 (1.01)	1 (1)	.40
GE2	2.68 (1.16)	3 (2)	2.91 (0.98)	3 (2)	.05
GE3	0.60 (1.17)	0 (1)	0.44 (0.77)	0 (1)	.11
GE4	1.19 (1.03)	1 (2)	1.12 (0.99)	1 (2)	.83
GE5	1.22 (1.14)	1 (2)	1.18 (1.05)	1 (1.5)	.70
GE6	1.42 (1.26)	1 (1)	1.32 (1.06)	1 (1)	.59
Functional Well-Being					
GF1	2.12 (1.22)	1 (1)	2.21 (1.20)	1 (1)	.23
GF2	2.30 (1.20)	2 (2)	2.32 (1.12)	2 (2)	.34
GF3	2.50 (1.10)	3 (1)	2.51 (1.11)	3 (1)	.81
GF4	2.58 (1.04)	3 (1)	2.55 (1.01)	3 (1)	.39
GF5	2.40 (1.18)	3 (1)	2.41 (1.15)	3 (1)	.25
GF6	2.53 (1.15)	3 (1)	2.65 (1.05)	3 (1)	.20
GF7	2.19 (1.15)	2 (1)	2.17 (1.07)	2 (1.5)	.81
Breast Cancer Subscale					
B1	0.75 (0.93)	0.25 (1)	0.72 (0.90)	0.5 (1)	.49
B2	0.58 (1.10)	0 (1)	0.52 (0.99)	0 (1)	.79
B3	0.68 (1.03)	0 (1)	0.63 (0.91)	0 (1)	>.99
B4	1.73 (1.04)	2 (1)	1.69 (1.01)	2 (1)	.81
B5	1.47 (1.41)	1 (2)	1.47 (1.38)	1 (2)	.49
B6	1.88 (1.41)	2 (2)	1.82 (1.45)	1.5 (2)	.06
B7	2.02 (1.38)	2 (2)	2.14 (1.33)	2 (2)	.33
B8	1.23 (1.34)	1 (2)	1.13 (1.37)	1 (2)	.11

Single items	Paper-based patient-reported outcomes		Electronic patient-reported outcomes		*P* value
	Mean (SD)	Median (Interquartile range)	Mean (SD)	Median (Interquartile range)	
B9	2.55 (1.18)	3 (1)	2.62 (1.17)	3 (2)	.58
P2	1.28 (1.11)	1 (2)	1.13 (1.05)	1 (1.75)	.18

[a]Statistically significant difference.

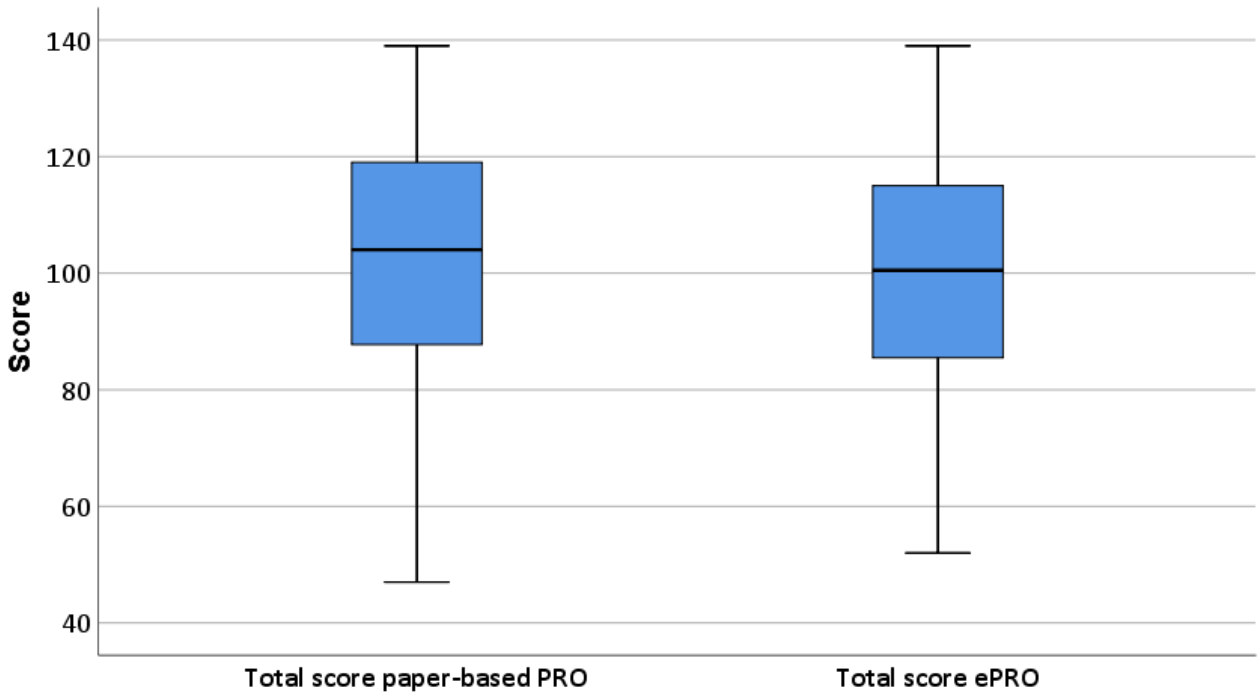
Table 4. Parallel forms reliability (Wilcoxon test) for scoring values of 5 Functional Assessment of Cancer Therapy-Breast (FACT-B) dimensions (subscales).

FACT-B[a] dimensions	FACT-B questions	Paper-based patient-reported outcomes		Electronic patient-reported outcomes		*P* value
		Mean (SD)	Median (Interquartile range)	Mean (SD)	Median (Interquartile range)	
Physical Well-Being Sum individual item scores	GP1-GP7	19.97 (6.11)	21.0 (9.0)	19.89 (5.88)	20.0 (8.25)	0.05
Social/Family Well-Being Sum individual item scores	GS1-GS7	22.88 (3.93)	24.0 (5.0)	22.34 (4.60)	23.33 (5.0)	0.25
Emotinal Well-Being Sum individual item scores	GE1-GE6	16.89 (4.84)	18.0 (6.0)	17.73 (4.68)	18.0 (6.0)	*.01*[b]
Functional Well-Being Sum individual item scores	GF1-GF7	16.73 (6.29)	18.0 (9.0)	16.75 (6.10)	18.0 (9.45)	0.69
Breast Cancer Subscale Sum individual item scores	B1-B9; P2	26.35 (6.28)	28.0 (9.0)	28.56 (7.11)	30.0 (11.33)	*<.001*[b]
FACT-B total score		102.66 (22.0)	106.33 (28.81)	104.39 (22.47)	107 (30.31)	0.05

[a]FACT-B: Functional Assessment of Cancer Therapy-Breast

[a]Statistically significant difference.

Figure 3. Distribution of total scores (Boxplot). PRO: patient-reported outcome.



Test of Internal Consistency

Table 5 shows the Spearman ρ correlation values and agreement rates, which were obtained for every question of the FACT-B questionnaire. In 27 questions, a high correlation (>.8) was found between paper-based PRO and ePRO, while the correlation was moderate (>.5) in the other 10 questions. In 20 questions, the correlation levels amounted to >.85. In all 37 correlated questions, agreement rates fluctuated between 65.7% and 91.8%.

Table 6 shows the results of internal consistency testing for the sum individual item subscale scores and the total FACT-B score between paper-based PRO and ePRO. In all functional (sub)scales, the rank correlation was moderate to high as Kendall tau coefficient ranged between.64 and.80 in the sum individual item subscale scores and amounted to.80 in the total FACT-B score. The analysis of internal consistency tests showed the highest correlation in the sum individual item subscale scores EWB and FWB. All results of internal consistency tests were statistically highly significant. For illustrative purposes, Figure 4 represents a strong positive correlation between the ePRO and paper-based total FACT-B scores. Each data point reflects individual scores of patients.

Table 5. Test of internal consistency in single items: results of correlation (Spearman ρ) and agreement analyses.

Dimensions	Spearman ρ	*P* value of Spearman ρ[a]	Agreement (%)
Physical Well-Being			
GP1	0.869	<.001	72.0
GP2	0.836	<.001	86.9
GP3	0.866	<.001	70.1
GP4	0.836	<.001	75.4
GP5	0.837	<.001	68.3
GP6	0.842	<.001	71.5
GP7	0.889	<.001	86.4
Social/Family Well-Being			
GS1	0.782	<.001	77.8
GS2	0.880	<.001	85.4
GS3	0.908	<.001	87.8
GS4	0.782	<.001	76.3
GS5	0.829	<.001	81.3
GS6	0.931	<.001	91.8
GS7	0.747	<.001	77.0
Emotional Well-Being			
GE1	0.753	<.001	71.0
GE2	0.525	<.001	66.6
GE3	0.796	<.001	82.3
GE4	0.868	<.001	78.0
GE5	0.931	<.001	85.8
GE6	0.733	<.001	65.7
Functional Well-Being			
GF1	0.881	<.001	74.0
GF2	0.770	<.001	69.7
GF3	0.889	<.001	81.0
GF4	0.821	<.001	71.5
GF5	0.934	<.001	87.2
GF6	0.910	<.001	82.4
GF7	0.897	<.001	75.2
Breast Cancer Subscale			
B1	0.858	<.001	86.4
B2	0.925	<.001	87.0
B3	0.904	<.001	87.9
B4	0.777	<.001	80.7
B5	0.900	<.001	80.6
B6	0.949	<.001	83.6
B7	0.843	<.001	69.5
B8	0.945	<.001	86.0
B9	0.684	<.001	70.1
P2	0.850	<.001	73.8

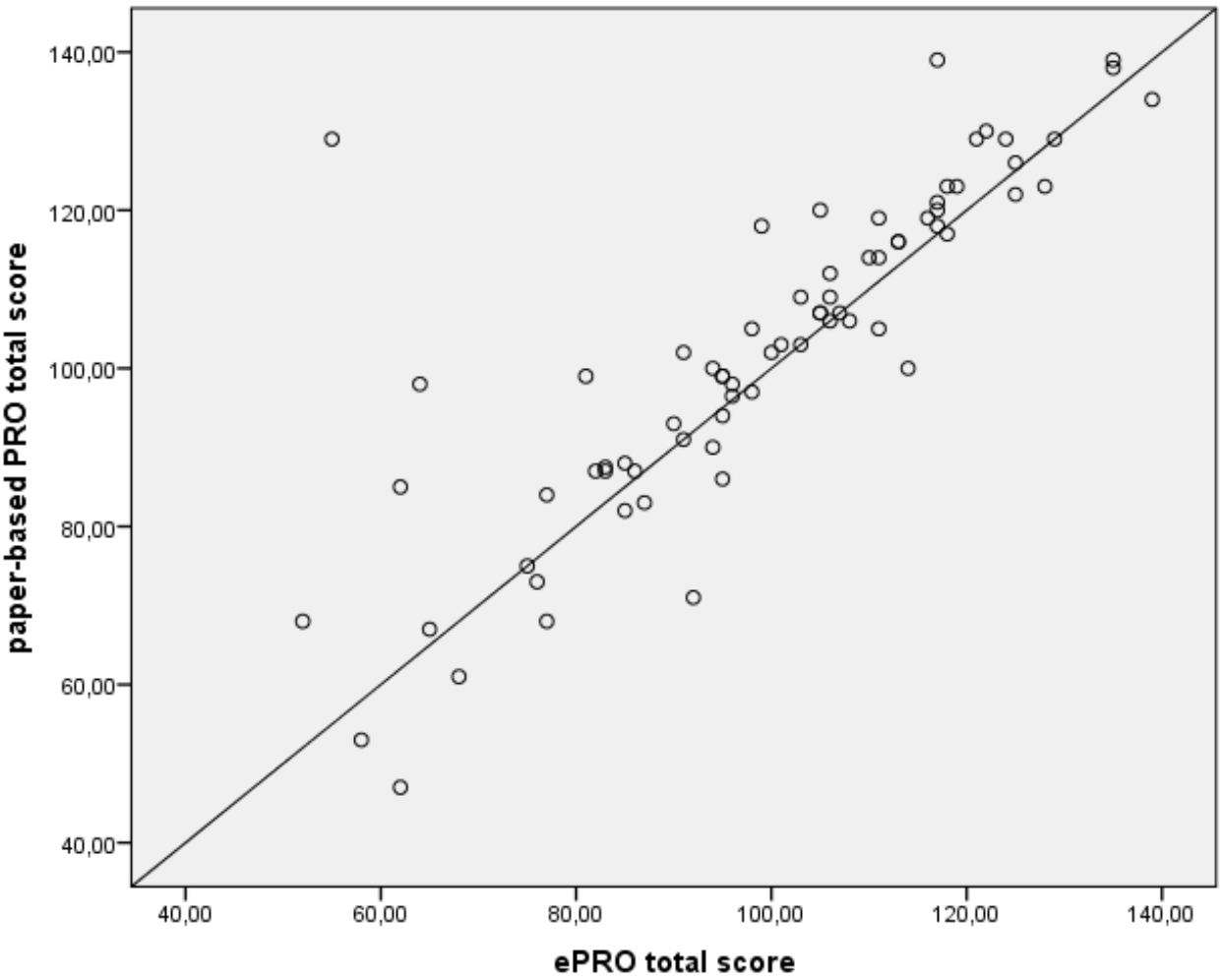
[a]Statistically highly significant correlations.

Table 6. Test of internal consistency in the individual subscale item scores and the total score: Kendall tau analysis.

Dimensions	Kendall tau (95% CI)	*P* value of Kendall tau[a]
Physical Well-Being Sum individual item scores	0.784 (0.723-0.835)	*<.001*
Social/Family Well-Being Sum individual item scores	0.648 (0.545-0.749)	*<.001*
Emotional Well-Being Sum individual item scores	0.737 (0.638-0.820)	*<.001*
Functional Well-Being Sum individual item scores	0.797 (0.731-0.858)	*<.001*
Breast Cancer Subscale Sum individual item scores	0.638 (0.536-0.724)	*<.001*
Total score	0.801 (0.741-0.852)	*<.001*

[a]Statistically significant correlations.

Figure 4. Correlation between electronic patient-reported outcome (ePRO) and paper-based total Functional Assessment of Cancer Therapy-Breast scores.



Discussion

Principal Results

In both dimensions of reliability (parallel forms reliability and internal consistency), we found high correlations when comparing single items in the patients' response behavior between paper-based PRO and ePRO in the FACT-B questionnaire. In the test of parallel forms reliability, we only found statistically significant differences in all but 2 questions. In the test of consistency, moderate to high correlations were found in all 37 single items and all sum individual item subscale scores. Based on the empirical results, the PiiA tool's ePRO version of FACT-B seems to be reliable for measuring the HRQoL in patients with metastatic and adjuvant breast cancer. According to these results, we would not expect that future use of the PiiA tool in the *same* patient group will show marked differences between the paper-based PRO and ePRO version

of FACT-B or that patients' response behavior is significantly influenced by the survey tool after transfer to electronically based patient surveys. Therefore, the tool is suitable for determining the HRQoL in patients with metastatic or adjuvant breast cancer.

Comparison With Prior Work

Although ePRO apps are being used increasingly frequently, paper-based surveys of PRO still predominate in clinical research because reliable, electronically validated questionnaires are lacking. One of the most commonly used questionnaires for measuring the HRQoL, especially in patients with breast cancer, is the FACT-B, for which there is a reliable, paper-based format in many languages but no reliable, electronically based version exists in German. Using the electronically based version of FACT-B and other PRO without verifying the reliability could endanger the significance of ePRO surveys. Thus, a corresponding analysis in relation to differences and correlations between the paper-based version of FACT-B and newly developed Web-based tools are of great importance. It can be assumed that many aspects (ie, sociodemographic aspects, technical skills, health condition, and, perhaps, design specifics of the ePRO tool) may influence both patients' willingness to use the tool and their response behavior, which underlines the need for reliability analyses [10,24,35-37]. However, almost no scientific studies have dealt with the reliability of ePRO questionnaires. Now, our working group has verified the reliability of an ePRO version of the FACT-B questionnaire in this paper, as well as the reliability of a tool for the ePRO measurement of the HRQoL in the EORTC questionnaire [40].

Limitations

Despite positive results, some limitations of the study design and methodological implementation should be mentioned, which could reduce the representativeness of the data. In 2 of 37 questions, we found statistically significant differences by the Wilcoxon test (parallel forms reliability). This can be a random observation based on numerous tests performed. A further explanation could be that patients were less concentrated owing to time pressure. Patients (both in arms A and B) were required to complete both paper-based and ePRO during an outpatient hospital visit. Patients were surveyed while they were receiving chemotherapy and were not permitted to take the questionnaire home to complete it; this also explains increasing missing values, especially in the last quarter of both questionnaire versions; possibly the length of the survey (paper-based and ePRO FACT-B and EORTC QLQ C-30, socioeconomic data, and evaluation of the tool) was too extensive for some patients. Possibly, the burden of disease and the therapy were potentially affecting the ability of some patients to complete both the paper- and tablet-based version of the questionnaire during an outpatient visit. Hence, a possible limiting factor was an inadequate screening as to whether all patients were able to cope with the psychological burden of participating in the study, as it is known that psycho-oncological distress is a commonly associated burden that could potentially influence the willingness to use ePRO and as a result ePRO's reliability [37,44]. Nonetheless, the influence of psycho-oncological stress was likely low in this study, as the test of internal consistency showed no abnormalities. Both the individual questions and the sum individual item subscale scores showed consistently statistically significant correlations. It also needs to be noted that there was possibly a selection bias, as we did not examine whether the HRQoL was lower and the psychological distress was higher in those patients who could not be motivated to participate in the study. Therefore, it cannot be claimed with certainty that the tool *per se* is reliable for all patients with metastatic and adjuvant breast cancer; hence, further studies are needed that focus on the willingness to use ePRO depending on the state of health [10].

Strengths of the Study

Although ePRO is being used more and more often, questionnaires with proven reliability and validity are lacking. FACT-B is one of the most commonly used questionnaires worldwide for measuring the HRQoL in patients with breast cancer, but a reliable ePRO version is also missing here. One of the strengths of this study is that the reliability of a new tool for the ePRO measurement in patients with breast cancer was analyzed, while other studies often assign paper-based versions to a tablet-based format without verifying the reliability. The reliability of the PiiA tool could be ascertained for the questionnaires FACT-B (this paper) and EORTC QLQ C-30 [40] for measuring the HRQoL in metastatic and adjuvant breast cancer patients. The second strength is the methodological approach of the study ePROCOM and the statistical evaluation as all patients completed both the paper-based and ePRO questionnaire, and the reliability was ascertained in a multidimensional fashion (parallel forms reliability as well as internal consistency). Finally, the third strength points to the fact that ePRO tools are reliable and suitable in both the adjuvant and metastatic settings, although hurdles can be expected in these patient groups, in particular, depending on their health status, HRQoL, sociodemographic specifics, and technical ability and experience [10,37]. The results of this study can improve the quality of ePRO measurements as they seem to be generalizable, and the PiiA app of FACT-B can be used for reliable e-based measurement of the HRQoL in other studies and clinical routine.

Conclusions

Electronically based PRO is constantly being adopted in clinical research and clinical routine, which underlines the need for reliable questionnaires. The evaluated electronic version of the FACT-B is reliable for patients with breast cancer in an adjuvant or metastatic setting because high correlations were found in almost all questions, all subscales, and the total score. Thus, this study concludes that the validated paper-based questionnaire of FACT-B and the ePRO tool are equal.

Acknowledgments

This study was supported by the Klaus Tschira Stiftung, gemeinnützige GmbH.

Conflicts of Interest

PG has received honoraria from Novartis and financial support for symposia from Novartis, Roche, and PharmaMar.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 161KB - jmir_v21i1e10004_app1.pdf]

References

1. Robert Koch Institute. Berlin. 2016. Bericht zum Krebsgeschehen in Deutschland 2016 URL: http://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebsgeschehen/Krebsgeschehen_download.pdf?__blob=publicationFile [WebCite Cache ID 75BH78Gjb]
2. Verma S, Miles D, Gianni L, Krop IE, Welslau M, Baselga J, EMILIA Study Group. Trastuzumab emtansine for HER2-positive advanced breast cancer. N Engl J Med 2012 Nov 08;367(19):1783-1791 [FREE Full text] [doi: 10.1056/NEJMoa1209124] [Medline: 23020162]
3. Heil J, Gondos A, Rauch G, Marmé F, Rom J, Golatta M, et al. Outcome analysis of patients with primary breast cancer initially treated at a certified academic breast unit. Breast 2012 Jun;21(3):303-308. [doi: 10.1016/j.breast.2012.01.009] [Medline: 22310244]
4. Dufresne A, Pivot X, Tournigand C, Facchini T, Altweegg T, Chaigneau L, et al. Impact of chemotherapy beyond the first line in patients with metastatic breast cancer. Breast Cancer Res Treat 2008 Jan;107(2):275-279. [doi: 10.1007/s10549-007-9550-7] [Medline: 17380382]
5. Kassam F, Enright K, Dent R, Dranitsaris G, Myers J, Flynn C, et al. Survival outcomes for patients with metastatic triple-negative breast cancer: implications for clinical practice and trial design. Clin Breast Cancer 2009 Feb;9(1):29-33. [doi: 10.3816/CBC.2009.n.005] [Medline: 19299237]
6. Cortés J, Baselga J, Im Y, Im S, Pivot X, Ross G, et al. Health-related quality-of-life assessment in CLEOPATRA, a phase III study combining pertuzumab with trastuzumab and docetaxel in metastatic breast cancer. Ann Oncol 2013 Oct;24(10):2630-2635. [doi: 10.1093/annonc/mdt274] [Medline: 23868905]
7. Wallwiener M, Heindl F, Brucker SY, Taran F, Hartkopf A, Overkamp F, et al. Implementation and Feasibility of Electronic Patient-Reported Outcome (ePRO) Data Entry in the PRAEGNANT Real-Time Advanced and Metastatic Breast Cancer Registry. Geburtshilfe Frauenheilkd 2017 Aug;77(8):870-878 [FREE Full text] [doi: 10.1055/s-0043-116223] [Medline: 28845051]
8. Karamouzis MV, Ioannidis G, Rigatos G. Quality of life in metastatic breast cancer patients under chemotherapy or supportive care: a single-institution comparative study. Eur J Cancer Care (Engl) 2007 Sep;16(5):433-438. [doi: 10.1111/j.1365-2354.2006.00771.x] [Medline: 17760930]
9. Bitton A, Onega T, Tosteson ANA, Haas JS. Toward a better understanding of patient-reported outcomes in clinical practice. Am J Manag Care 2014 Apr;20(4):281-283 [FREE Full text] [Medline: 24884859]
10. Graf J, Simoes E, Wißlicen K, Rava L, Walter CB, Hartkopf A, et al. Willingness of Patients with Breast Cancer in the Adjuvant and Metastatic Setting to Use Electronic Surveys (ePRO) Depends on Sociodemographic Factors, Health-related Quality of Life, Disease Status and Computer Skills. Geburtshilfe Frauenheilkd 2016 May;76(5):535-541 [FREE Full text] [doi: 10.1055/s-0042-105872] [Medline: 27239062]
11. Reilly CM, Bruner DW, Mitchell SA, Minasian LM, Basch E, Dueck AC, et al. A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment. Support Care Cancer 2013 Jun;21(6):1525-1550 [FREE Full text] [doi: 10.1007/s00520-012-1688-0] [Medline: 23314601]
12. Fromme EK, Eilers KM, Mori M, Hsieh Y, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. J Clin Oncol 2004 Sep 01;22(17):3485-3490. [doi: 10.1200/JCO.2004.03.025] [Medline: 15337796]
13. Pakhomov SV, Jacobsen SJ, Chute CG, Roger VL. Agreement between patient-reported symptoms and their documentation in the medical record. Am J Manag Care 2008 Aug;14(8):530-539 [FREE Full text] [Medline: 18690769]
14. Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst 2009 Dec 02;101(23):1624-1632 [FREE Full text] [doi: 10.1093/jnci/djp386] [Medline: 19920223]
15. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-reported outcomes (PROs): putting the patient perspective in patient-centered outcomes research. Med Care 2013 Aug;51(8 Suppl 3):S73-S79 [FREE Full text] [doi: 10.1097/MLR.0b013e31829b1d84] [Medline: 23774513]

16. Valderas JM, Alonso J. Patient reported outcome measures: a model-based classification system for research and clinical practice. Qual Life Res 2008 Nov;17(9):1125-1135. [doi: 10.1007/s11136-008-9396-4] [Medline: 18836850]
17. Basch E. New frontiers in patient-reported outcomes: adverse event reporting, comparative effectiveness, and quality assessment. Annu Rev Med 2014;65:307-317. [doi: 10.1146/annurev-med-010713-141500] [Medline: 24274179]
18. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
19. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724. [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]
20. Wittmann V, Látos M, Horváth Z, Simonka Z, Paszt A, Lázár G, et al. What contributes to long-term quality of life in breast cancer patients who are undergoing surgery? Results of a multidimensional study. Qual Life Res 2017 Dec;26(8):2189-2199. [doi: 10.1007/s11136-017-1563-z] [Medline: 28357681]
21. Kreienberg R, Albert U, Follmann M, Kopp I, Kühn T, Wöckel A, et al. Leitlinienprogramm Onkologie. 2012. Interdisziplinäre S3-Leitlinie für die Diagnostik, Therapie und Nachsorge des Mammakarzinoms URL: https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/LL_Mammakarzinom_Langversion_4.0.pdf [accessed 2018-11-15] [WebCite Cache ID 73xoonywK]
22. Jones SE. Metastatic breast cancer: the treatment challenge. Clin Breast Cancer 2008 Jun;8(3):224-233. [doi: 10.3816/CBC.2008.n.025] [Medline: 18650152]
23. Stefanovic S, Wallwiener M, Karic U, Domschke C, Katic L, Taran F, et al. Patient-reported outcomes (PRO) focused on adverse events (PRO-AEs) in adjuvant and metastatic breast cancer: clinical and translational implications. Support Care Cancer 2017 Dec;25(2):549-558. [doi: 10.1007/s00520-016-3437-2] [Medline: 27747478]
24. Wallwiener M, Simoes E, Sokolov AN, Brucker SY, Fasching PA, Graf J. Health-related Quality of Life in Metastatic and Adjuvant Breast Cancer Patients. Geburtshilfe Frauenheilkd 2016 Oct;76(10):1065-1073 [FREE Full text] [doi: 10.1055/s-0042-113188] [Medline: 27761027]
25. Hahn E, Segawa E, Kaiser K, Cella D, Smith B. Validation of the Functional Assessment of Cancer Therapy-Breast (FACT-B-B) quality of life instrument. Journal of Clinical Oncology 2015;33(suppl e1):7753. [doi: 10.1200/jco.2015.33.15_suppl.e17753]
26. Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. J Clin Oncol 1997 Mar;15(3):974-986. [doi: 10.1200/JCO.1997.15.3.974] [Medline: 9060536]
27. FACIT Measurement System. FACT-B scoring template. In: FACIT Functional Assessment of Chronic Illness Therapy. Elmhurst: Selfpublishing; 2013.
28. Basch E, Goldfarb S. Electronic patient-reported outcomes for collecting sensitive information from patients. J Support Oncol 2009;7(3):98-99. [Medline: 19507457]
29. Abernethy AP, Herndon JE, Wheeler JL, Day JM, Hood L, Patwardhan M, et al. Feasibility and acceptability to patients of a longitudinal system for evaluating cancer-related symptoms and quality of life: pilot study of an e/Tablet data-collection system in academic oncology. J Pain Symptom Manage 2009 Jun;37(6):1027-1038. [doi: 10.1016/j.jpainsymman.2008.07.011] [Medline: 19394793]
30. Jones JB, Snyder CF, Wu AW. Issues in the design of Internet-based systems for collecting patient-reported outcomes. Qual Life Res 2007 Oct;16(8):1407-1417. [doi: 10.1007/s11136-007-9235-z] [Medline: 17668293]
31. Basch E, Artz D, Iasonos A, Speakman J, Shannon K, Lin K, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. J Am Med Inform Assoc 2007;14(3):264-268 [FREE Full text] [doi: 10.1197/jamia.M2177] [Medline: 17329732]
32. Basch E, Iasonos A, Barz A, Culkin A, Kris MG, Artz D, et al. Long-term toxicity monitoring via electronic patient-reported outcomes in patients receiving chemotherapy. J Clin Oncol 2007 Dec 01;25(34):5374-5380. [doi: 10.1200/JCO.2007.11.2243] [Medline: 18048818]
33. Snyder CF, Blackford AL, Wolff AC, Carducci MA, Herman JM, Wu AW, PatientViewpoint Scientific Advisory Board. Feasibility and value of PatientViewpoint: a web system for patient-reported outcomes assessment in clinical practice. Psychooncology 2013 Apr;22(4):895-901 [FREE Full text] [doi: 10.1002/pon.3087] [Medline: 22544513]
34. Trautmann F, Hentschel L, Hornemann B, Rentsch A, Baumann M, Ehninger G, et al. Electronic real-time assessment of patient-reported outcomes in routine care-first findings and experiences from the implementation in a comprehensive cancer center. Support Care Cancer 2016 Dec;24(7):3047-3056. [doi: 10.1007/s00520-016-3127-0] [Medline: 26887586]
35. Graf J, Moreno B, Wallwiener M, Menzel K, Brucker SY, Simoes E. [Practicability and Efficiency of E-health Applications in Patient-Reported Outcomes: State of and Need for Research]. Gesundheitswesen 2018 Dec 27. [doi: 10.1055/s-0044-100037] [Medline: 29486499]
36. Barentsz MW, Wessels H, van Diest PJ, Pijnappel RM, Haaring C, van der Pol CC, et al. Tablet, web-based, or paper questionnaires for measuring anxiety in patients suspected of breast cancer: patients' preferences and quality of collected data. J Med Internet Res 2014 Oct 31;16(10):e239 [FREE Full text] [doi: 10.2196/jmir.3578] [Medline: 25364951]

37. Hartkopf AD, Graf J, Simoes E, Keilmann L, Sickenberger N, Gass P, et al. Electronic-Based Patient-Reported Outcomes: Willingness, Needs, and Barriers in Adjuvant and Metastatic Breast Cancer Patients. JMIR Cancer 2017 Aug 07;3(2):e11 [FREE Full text] [doi: 10.2196/cancer.6996] [Medline: 28784595]
38. Campbell N, Ali F, Finlay AY, Salek SS. Equivalence of electronic and paper-based patient-reported outcome measures. Qual Life Res 2015 Aug;24(8):1949-1961. [doi: 10.1007/s11136-015-0937-3] [Medline: 25702266]
39. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, ISPOR ePRO Task Force. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009 Jun;12(4):419-429 [FREE Full text] [doi: 10.1111/j.1524-4733.2008.00470.x] [Medline: 19900250]
40. Wallwiener M, Matthies L, Simoes E, Keilmann L, Hartkopf AD, Sokolov AN, et al. Reliability of an e-PRO Tool of EORTC QLQ-C30 for Measurement of Health-Related Quality of Life in Patients With Breast Cancer: Prospective Randomized Trial. J Med Internet Res 2017 Sep 14;19(9):e322 [FREE Full text] [doi: 10.2196/jmir.8210] [Medline: 28912116]
41. Lachin JM. Statistical properties of randomization in clinical trials. Control Clin Trials 1988 Dec;9(4):289-311. [Medline: 3060315]
42. Schulz KF. Randomized controlled trials. Clin Obstet Gynecol 1998 Jun;41(2):245-256. [Medline: 9646957]
43. Kemmler G, Holzner B, Kopp M, Dünser M, Margreiter R, Greil R, et al. Comparison of two quality-of-life instruments for cancer patients: the functional assessment of cancer therapy-general and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30. J Clin Oncol 1999 Sep;17(9):2932-2940. [doi: 10.1200/JCO.1999.17.9.2932] [Medline: 10561373]
44. Schaeffeler N, Pfeiffer K, Ringwald J, Brucker S, Wallwiener M, Zipfel S, et al. Assessing the need for psychooncological support: screening instruments in combination with patients' subjective evaluation may define psychooncological pathways. Psychooncology 2015 Dec;24(12):1784-1791. [doi: 10.1002/pon.3855] [Medline: 26042392]

Abbreviations

BCS: Breast Cancer Subscale
ePROCOM: electronic Patient-Reported Outcomes and Compliance Analysis
EORTC: European Organisation for Research and Treatment of Cancer
EWB: Emotional Well-Being
FACT-B: Functional Assessment of Cancer Therapy-Breast
FWB: Functional Well-Being
HRQoL: health-related quality of life
PEPPER: Patient Engagement Pilotstudie Mammakarzinom - individualisierte und Ressourcen-effiziente Patient Reported Outcomes Erfassung durch Digitale Therapieunterstützungssysteme
PiiA: Patient-informiert-interaktiv-Arzt
PRO: patient-reported outcomes
PWB: Physical Well-Being
QLQ: quality of life questionnaire
SWB: Social/Family Well-Being

Edited by G Eysenbach; submitted 31.01.18; peer-reviewed by A Birukov, Z Zhou; comments to author 25.08.18; revised version received 13.10.18; accepted 13.10.18; published 22.01.19.

Please cite as:
Matthies LM, Taran FA, Keilmann L, Schneeweiss A, Simoes E, Hartkopf AD, Sokolov AN, Walter CB, Sickenberger N, Wallwiener S, Feisst M, Gass P, Lux MP, Schuetz F, Fasching PA, Sohn C, Brucker SY, Graf J, Wallwiener M
An Electronic Patient-Reported Outcome Tool for the FACT-B (Functional Assessment of Cancer Therapy-Breast) Questionnaire for Measuring the Health-Related Quality of Life in Patients With Breast Cancer: Reliability Study
J Med Internet Res 2019;21(1):e10004
URL: http://www.jmir.org/2019/1/e10004/
doi:10.2196/10004
PMID:30668517

Original Paper

The Most Influential Medical Journals According to Wikipedia: Quantitative Analysis

Dariusz Jemielniak[1*], PhD; Gwinyai Masukume[2*], MB BCh, MSc; Maciej Wilamowski[3*], MS, PhD

[1]Department of Management in Networked and Digital Societies, Kozminski University, Warszawa, Poland

[2]The Irish Centre for Fetal and Neonatal Translational Research, Department of Obstetrics and Gynaecology, University College Cork, Cork, Ireland

[3]Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland

[*]all authors contributed equally

Corresponding Author:
Dariusz Jemielniak, PhD
Department of Management in Networked and Digital Societies
Kozminski University
Jagiellonska 59
Warszawa, 03301
Poland
Phone: 48 604901352
Email: darekj@kozminski.edu.pl

Abstract

Background: Wikipedia, the multilingual encyclopedia, was founded in 2001 and is the world's largest and most visited online general reference website. It is widely used by health care professionals and students. The inclusion of journal articles in Wikipedia is of scholarly interest, but the time taken for a journal article to be included in Wikipedia, from the moment of its publication to its incorporation into Wikipedia, is unclear.

Objective: We aimed to determine the ranking of the most cited journals by their representation in the English-language medical pages of Wikipedia. In addition, we evaluated the number of days between publication of journal articles and their citation in Wikipedia medical pages, treating this measure as a proxy for the information-diffusion rate.

Methods: We retrieved the dates when articles were included in Wikipedia and the date of journal publication from Crossref by using an application programming interface.

Results: From 11,325 Wikipedia medical articles, we identified citations to 137,889 journal articles from over 15,000 journals. There was a large spike in the number of journal articles published in or after 2002 that were cited by Wikipedia. The higher the importance of a Wikipedia article, the higher was the mean number of journal citations it contained (top article, 48.13 [SD 33.67]; lowest article, 6.44 [SD 9.33]). However, the importance of the Wikipedia article did not affect the speed of reference addition. The *Cochrane Database of Systematic Reviews* was the most cited journal by Wikipedia, followed by *The New England Journal of Medicine* and *The Lancet*. The multidisciplinary journals *Nature*, *Science*, and the *Proceedings of the National Academy of Sciences* were among the top 10 journals with the highest Wikipedia medical article citations. For the top biomedical journal papers cited in Wikipedia's medical pages in 2016-2017, it took about 90 days (3 months) for the citation to be used in Wikipedia.

Conclusions: We found evidence of "recentism," which refers to preferential citation of recently published journal articles in Wikipedia. Traditional high-impact medical and multidisciplinary journals were extensively cited by Wikipedia, suggesting that Wikipedia medical articles have robust underpinnings. In keeping with the Wikipedia policy of citing reviews/secondary sources in preference to primary sources, the *Cochrane Database of Systematic Reviews* was the most referenced journal.

(J Med Internet Res 2019;21(1):e11429) doi:10.2196/11429

KEYWORDS

citizen science; medical journals; open knowledge; Wikipedia; knowledge translation; journalology; medical publishing; scholarly publishing

Introduction

Wikipedia, the multilingual encyclopedia, is the world's largest and most visited online general reference website and, arguably, the largest collaborative project of humankind [1]. Wikipedia reflects the state of scientific knowledge but also shapes science; ideas that are integrated into the encyclopedia are used more in scientific journals [2]. There is evidently a feedback loop between Wikipedia and scholarly journals, which accelerates research. Indeed, traditional journals are increasingly citing Wikipedia formally [3], and the general distrust of Wikipedia in academic circles is decreasing [4].

Wikipedia is widely used by health care professionals and students as well as educators, journalists, and policy makers, among others [5,6]. In fact, medical students perform better on tests when they use Wikipedia as compared to standard medical digital textbooks (statistically significant difference) or a contemporary point-of-care medical website (statistically nonsignificant difference) [7]. The quality of information on Wikipedia on medical topics is generally high [8,9], although this quality is, admittedly, culturally influenced [10] and partly dependent on the editor's experience [11], with varying article readability [12,13].

The inclusion of academic articles into Wikipedia remains a topic of interest for scholars from various fields [14]. However, it is still not entirely clear how often and quickly recent sources are used to support Wikipedia's medical articles, from the point of publication to their incorporation into Wikipedia [15], and previous studies on the topic are limited in scope [16].

In this study, we analyzed 39,564 medical articles from the English-language Wikipedia to determine the time taken for journal publications to reach Wikipedia and to identify journal outlets that are most likely to be included in Wikipedia. We developed a ranking for medical journals based on their representation on Wikipedia. We hypothesized that the time taken for journal publications to reach Wikipedia was declining and that high-impact factor journals were represented more often.

Methods

We analyzed the number of days between journal article publication and its citation in the English-language Wikipedia, treating this measure as a proxy for the information-diffusion rate. For our analysis, we selected 39,561 medical articles on Wikipedia that were marked as "medical" by the Wikipedia community [17] (as of October 10th, 2017).

It is worth mentioning that not all articles tagged as medical (~25%) were expected to cite the scholarly literature because of the topics they covered or the early stage of their development. Among other Wikipedia special pages, redirect (~9%), category (~9%), and template pages (2%) were not expected to have citations [17].

From 11,314 Wikipedia medical articles, we found citations to 137,889 articles from over 15,000 journals. We retrieved the dates on which the articles were added to Wikipedia and the date of publication from Crossref for 108,600 references using an application programming interface. In 8,384 of these references, the date of addition to Wikipedia preceded the official date of publication, which does not necessarily signify an error, but might refer to preprints (Multimedia Appendix 1).

Results

When we analyzed the publication dates in the citations, we observed two important points. First, there was a clear increase in the number of articles published in or after 2002. Since Wikipedia was started in 2001, this finding is not surprising. Wikipedia editors have a clear preference for adding new sources (Figure 1), which is expected because "recentism" is an established phenomenon on Wikipedia, wherein coverage of recent events is disproportionately greater, and the Wikipedia community itself considers it a factor to be accounted for [18]. In the case of medical research, there is obvious value in focusing on more recent studies, as older studies may be obsolete. Nevertheless, the historical long view tends to be lost consequently. Second, we grouped the articles according to their importance decided by the Wikipedia community [17] and calculated the average number of citations per article (Table 1). An example of an article of highest (top) importance would be that on cancer (crucial to medicine). An abdominal pain article (directly affects many readers) would be of high importance; an abdominal mass article (interesting to many readers) would be of mid-importance, and an article on McBurney's point, an anatomical point in the abdomen, would be of low importance (other articles of low importance include hospitals, very rare diseases, and individuals).

As expected, the higher the importance of an article, the more citations it contained on an average. For further analysis, we clustered articles of "top", "high", and "medium" importance together and considered them as "important" articles, because their total count was roughly similar to the number of articles of "low" importance. Only about half of the articles had citations to journal articles, which is consistent with previous research [5].

We had two further working hypotheses. First, we assumed that highly important articles would be updated more often, and there would be a higher priority to update the references with the most recent research results. In the first two histograms in Figure 2, the blue line that represents the number of references for articles of higher importance is above the green line. However, in the third histogram both lines almost overlap, suggesting that before 2012, there was a higher priority for articles of higher importance. However, that was not the case after 2012, possibly due to equalization of maturity of articles in higher- and lower-importance groups or because this time period was not long enough for new research results that would require reference updating. Thus, after 2012, we did not observe a difference in the age distribution of the referenced articles. Second, we assumed that the more obscure articles are more likely to be developed by editors with a conflict of interest (COI), motivated to promote their own work on Wikipedia, who add references immediately upon their publication.

Figure 1. Journal articles cited in Wikipedia according to their year of publication.

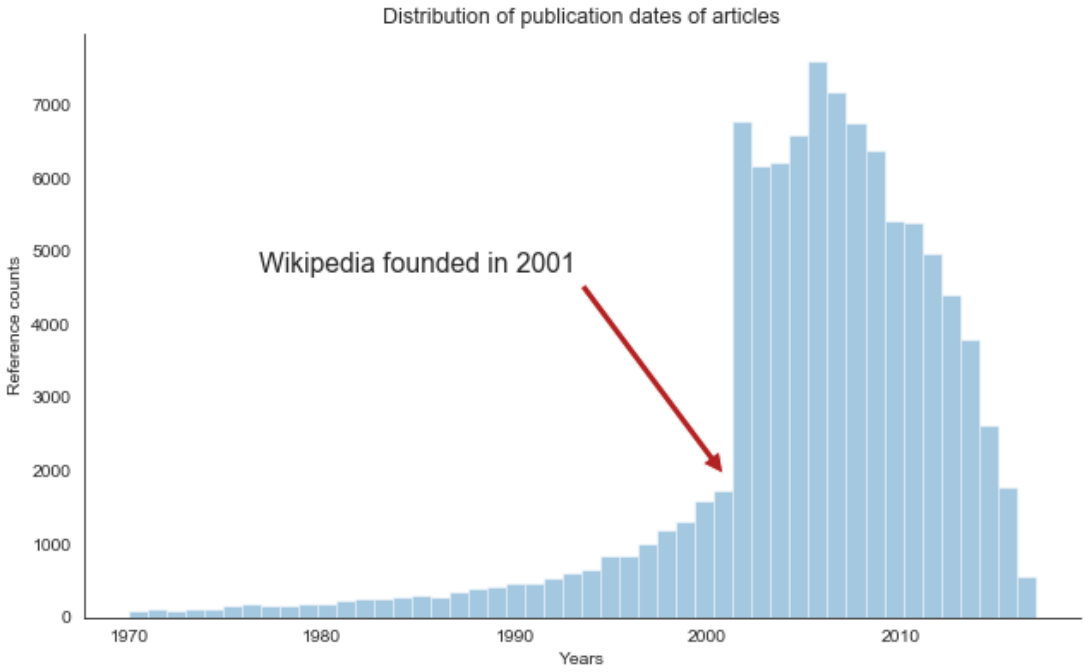
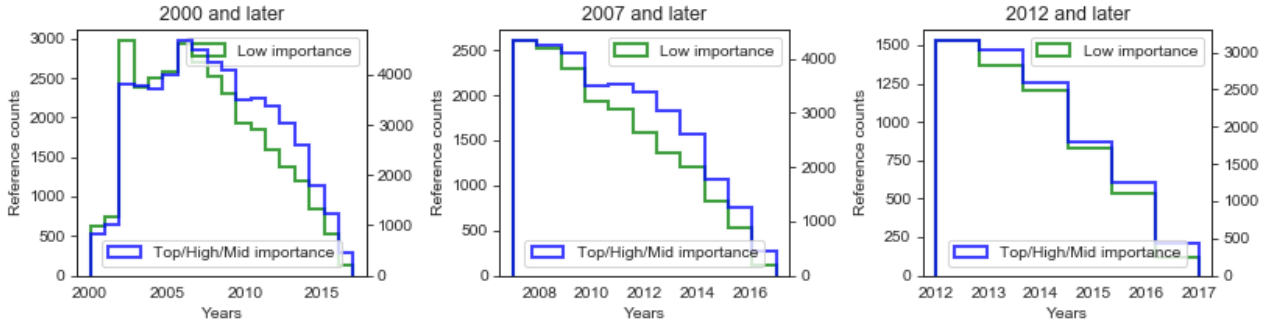


Table 1. The average number of references according to importance of the Wikipedia article.

Article importance	Articles, n (%)	References, mean (SD)	References, median (interquartile range)
Top	90 (0.8)	48.13 (33.67)	44.5 (17.25-67.5)
High	783 (6.9)	19.51 (22.21)	11 (5-26)
Medium	4532 (40.1)	9.39 (13.31)	5 (2-11)
Low	5905 (52.2)	6.44 (9.33)	3 (2-7)

Figure 2. Histograms of the number of citations between the two quality groups.



If the articles of lower importance were prone to quick referencing by editors with COI, the green histograms in Figure 2 would be much flatter, especially in the new articles (after 2012). For the medical sciences, this was an important observation, because Wikipedia is not solely focused on developing new articles and does not discriminate among topics: References are added equally to the most popular and important articles and the most obscure articles.

As observed, the average age of a referenced article does not have to be correlated with the number of days between the journal article publication date and the date of its citation on Wikipedia. To verify our second hypothesis, we analyzed the data in detail. First, we plotted histograms for both importance groups in three time periods (2012 and later, 2007 and later, 2000 and later; Figure 3). The distribution was smooth and remained independent of the quality and importance of articles on Wikipedia. Independent of the time perspective and without focusing on just the important articles, the distributions were similar between both importance groups (Figure 4). For easier comparison, we overlaid the estimated distributions in Figure 4. In all cases, the empirical distribution was best described by

a beta distribution. In every analyzed period, the estimated distribution had higher kurtosis for higher-importance articles. The differences in distributions may not be substantial, even if they are statistically significant. Thus, the time taken for breaking research to reflect on Wikipedia is a stable, reliable measure.

We thereafter analyzed the change in time between publication and citation on Wikipedia over time and the difference between the two importance groups (Figure 5). We found that the time from journal article publication to incorporation into Wikipedia has been declining substantially from 2001, when Wikipedia was started. Addition of the historical canon explains the lengthy time to incorporation during the first few years of Wikipedia's existence. In following years, as the canonical works were already covered, the time to incorporation in Wikipedia decreased. In 2016, the median time for articles of higher importance to be referenced in Wikipedia was 120 days; for articles of lower importance, 150 days; and for the 10 most highly cited journals, <90 days. As of mid-2018, it took about 3 months for articles published in high-impact journals to be incorporated into Wikipedia. Over time, with the median value calculated annually, it took up to 30% longer for articles of lower importance to be cited on Wikipedia as compared to articles of higher importance. Although in some years, there was little-to-no difference in the time to citation, the references of articles of higher importance in Wikipedia were updated faster than those of articles of lower importance.

Next, we analyzed the characteristics of the most cited journals of all analyzed articles. On comparing different periods, we observed relative stability in the rankings for most journals (Table 2), with two exceptions: *Proceedings of the National Academy of Sciences* (PNAS), whose position in the ranking dropped systematically over time, and *PLOS One*, which is characterized by a rising trend (Table 2). The change in the number of citations over time is clearly visible for PNAS (Figure 6). Furthermore, the number of citations for a given journal is characterized by spikes and sharp changes from year to year in some cases, which is only natural and may result from special issues on a topic, conferences coverage, or dedicated efforts of Wikipedia editors to cover specific topics. A list of the top 200 journal articles according to Wikipedia is presented in Multimedia Appendix 2.

Figure 3. Estimated distributions of the number of days between publication and citation on Wikipedia.

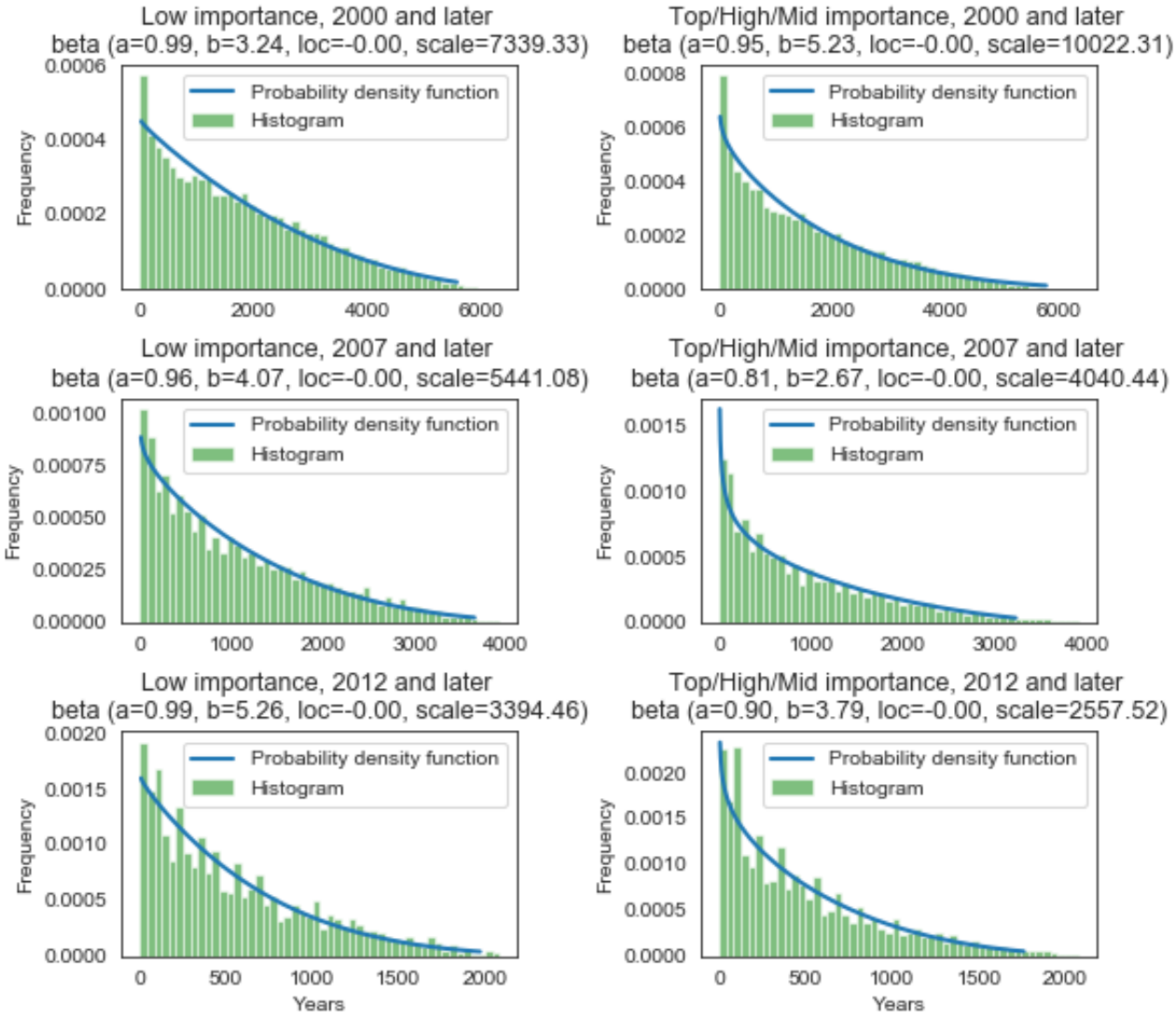


Figure 4. Comparison of the estimated distributions of the two quality groups.

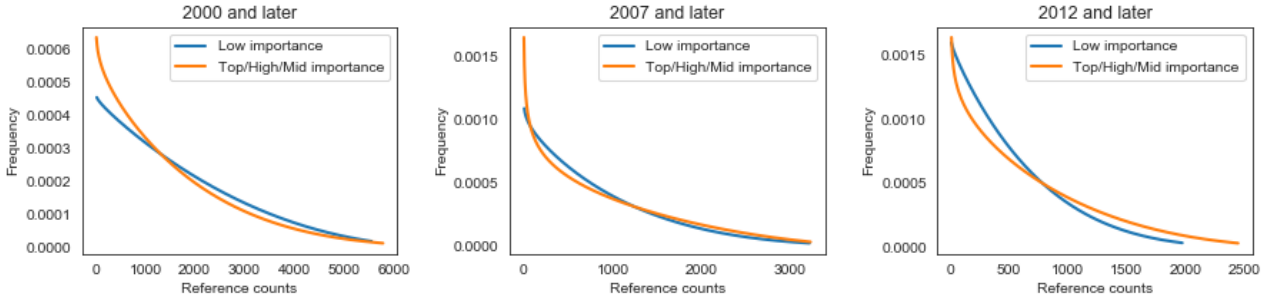
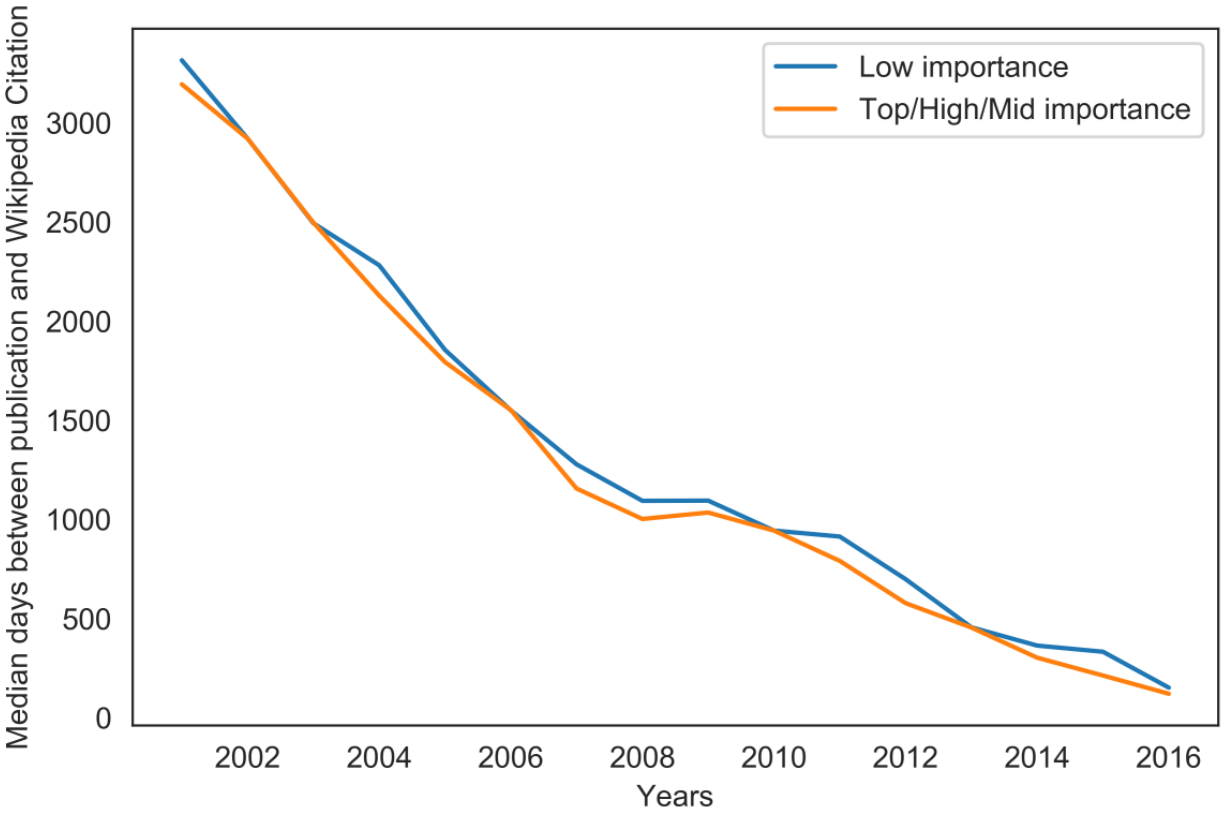


Figure 5. Change in average time between publication and Wikipedia citation over time.



To better understand the process of dissemination in Wikipedia, we calculated the time from publication to citation for individual journals (Figure 6). The data for the nine most cited journals show that there are some common characteristics in the journal specifics. In all cases, we observed clear spikes on day 0, as a lot of breaking medical research reflected on Wikipedia immediately, which confirms our previous findings. The next year shows a huge decline in the number of citations.

Cochrane Database of Systematic Reviews stands out as a very commonly cited source, with over 300 counts for day 0. Since this journal focuses on review articles and Wikipedia favors review articles over primary sources [19], this observation is not surprising. *PLOS One* stands out too, as it is very quickly reflected on Wikipedia. This multidisciplinary mega-journal was established in 2006; given that some articles require over 10 years to be reflected, the young age of *PLOS One* may have affected its peculiarity. Moreover, *PLOS One* is an immediate open-access journal, which is in line with many Wikipedia editors' philosophy. Interestingly, although PNAS is also a multidisciplinary journal, albeit not with fully open access, it does not rank among the top journals in our study. PNAS seemed to lose some of its importance among the editors over time.

Interestingly, *Nature* and *Science* either reach Wikipedia relatively slower or are more systematically backlogged (editors add journal articles from the further past). These journals do not focus on medicine specifically, which may explain a flatter distribution: In other words, publishing in these journals sooner or later results in coverage by Wikipedia. This is not necessarily the case for journals dedicated solely to medicine. Articles in medicine-only journals tend to appear on Wikipedia nearly immediately or are quite unlikely to ever be reflected.

Table 2. Top 10 most cited journals in three time periods (cumulative). The total number of citations in an analyzed period is presented within parentheses.

Rank	2000 and later	2007 and later	2012 and later
1	Cochrane Database of Systematic Reviews (1388)	Cochrane Database of Systematic Reviews (1073)	Cochrane Database of Systematic Reviews (720)
2	The New England Journal of Medicine (1291)	The New England Journal of Medicine (731)	PLOS One (352)
3	The Lancet (1156)	PLOS One (616)	The New England Journal of Medicine (255)
4	The BMJ (921)	The BMJ (557)	Annals of Internal Medicine (200)
5	Proceedings of the National Academy of Sciences of the United States of America (888)	The Lancet (424)	The Lancet (196)
6	JAMA: The Journal of the American Medical Association (822)	JAMA: The Journal of the American Medical Association (422)	The BMJ (192)
7	Nature (802)	Nature (357)	JAMA: The Journal of the American Medical Association (153)
8	Science (739)	Proceedings of the National Academy of Sciences of the United States of America (335)	Nature (147)
9	PLOS One (620)	Annals of Internal Medicine (273)	Neurology (110)
10	Journal of Biological Chemistry (560)	Neurology (224)	Proceedings of the National Academy of Sciences of the United States of America (94)

Our final finding is based on a ranking of the top 60 journals most commonly cited in Wikipedia medical articles. The differences between our ranking and those of Journal Citation Reports (JCR) or Scientific Journal Ranking (SJR) would indicate the specifics of Wikipedia (what is likely to be cited by a medical journal can differ from what is typically cited on Wikipedia; for instance, reviews and meta-analyses are more likely to be reflected in Wikipedia). However, a direct comparison is difficult, as other rankings have different criteria for considering a journal as medical, and general science journals would have to be included in some way.

The top 30 most commonly cited journals are much more diverse in the number of citations, and thus, we decided to offer a ranking. However, for positions 30 to 60, we did not rank the positions and listed the journals alphabetically, as the differences among them were too small.

As noted above, changes in the number of Wikipedia citations are dynamic over time, including irregular spikes and visible periods of maturation after Wikipedia's inception. Therefore, we propose a score for ranking, which accommodates the fast-changing environment of Wikipedia.

We propose a score calculated as an exponential moving average of a year-to-year ranking based on the number of citations. Using ranking instead of the absolute number of citations smoothens out the spikes and influx of citations that followed the creation of Wikipedia (Figure 7). Next, a moving average smoothens out large yearly variability in the ranking position. We believe this construction of ranking fairly reflects the relative popularity of journals in Wikipedia.

Our ranking for the top 30 journals is proposed in Table 3. There is a lot of variability for journals in positions above 30. Interestingly, only a few journals were considered to be best as per Wikipedia. Our ranking has an advantage over JCR or SJR rankings: It addresses the problem of general journals receiving high values for articles that are not related to medicine, as only medicine-related citations matter in our ranking.

Figure 6. The number of days from journal article publication to Wikipedia citation for the top nine journals. Each graph includes the name of the journal, the year the journal was founded, and the type of journal access at the top.

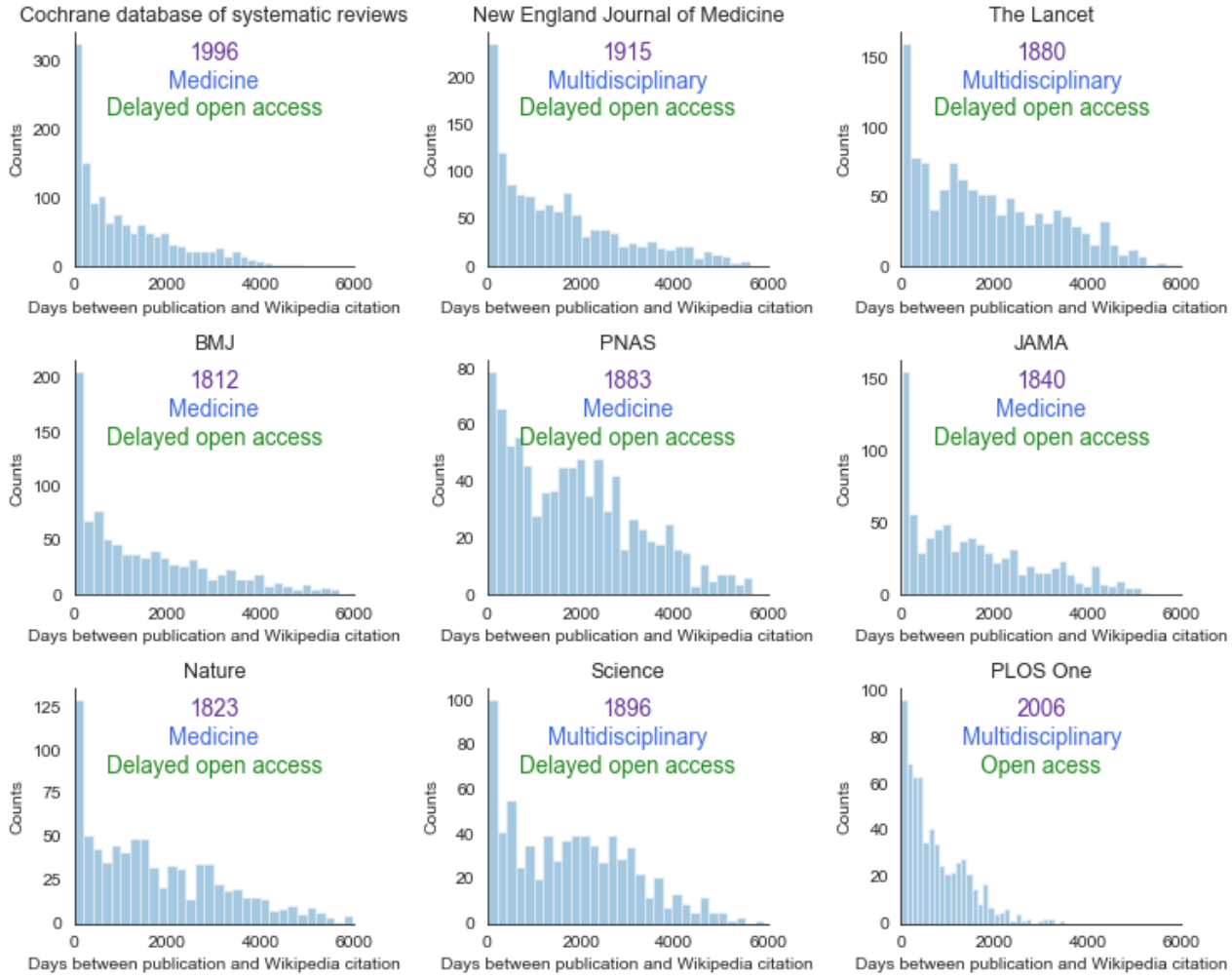


Figure 7. Journal citations over time.

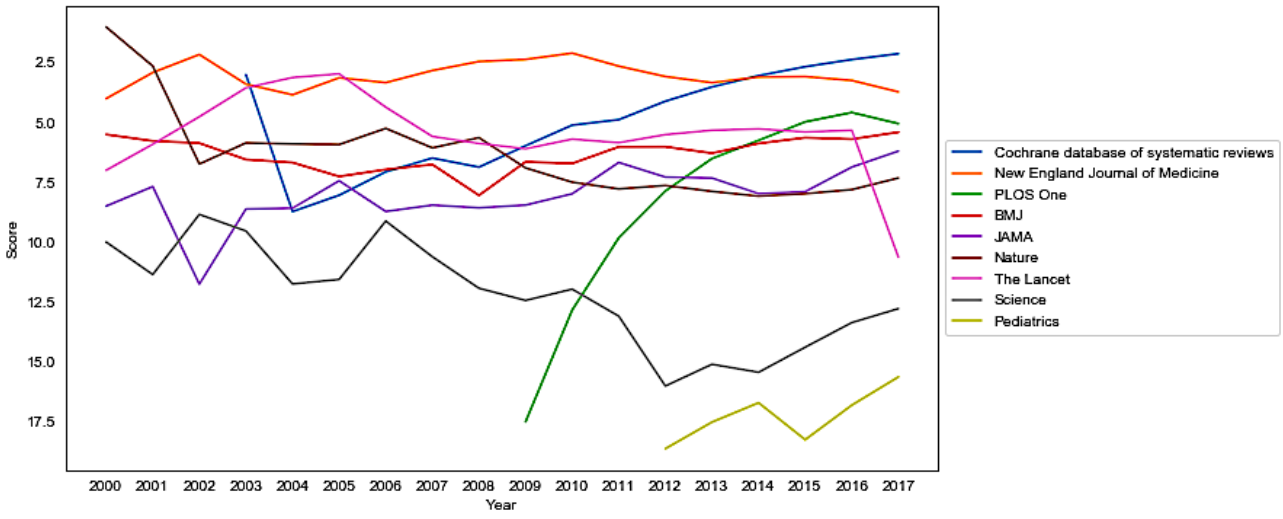


Table 3. Top 30 journals according to the study results.

Journal	Rank
The Cochrane Database of Systematic Reviews	1
The New England Journal of Medicine	2
PLOS One	3
The BMJ	4
JAMA: The Journal of the American Medical Association	5
Nature	6
The Lancet	7
Science	8
Pediatrics	9
Blood	10
Annals of Internal Medicine	11
Circulation	12
Neurology	13
BMJ Open	14
Journal of Clinical Oncology	15
Scientific Reports	16
Clinical Infectious Diseases	17
The Journal of Clinical Psychiatry	18
Nature Communications	19
Stroke	20
Emerging Infectious Diseases	21
Biochemical and Biophysical Research Communications	22
Neuropharmacology	23
Canadian Medical Association Journal	24
The Lancet Infectious Diseases	25
The American Journal of Medicine	26
Gastroenterology	27
International Journal of Cancer	28
Academic Emergency Medicine	29
Archives of Disease in Childhood	30

Discussion

We found evidence of "recentism," which refers to preferential citation of recently published journal articles in Wikipedia. Traditional high-impact medical and multidisciplinary journals were highly cited by Wikipedia, suggesting that Wikipedia medical articles have robust underpinnings. In keeping with the Wikipedia policy of citing reviews/secondary sources over primary sources, the *Cochrane Database of Systematic Reviews* was the most referenced journal, possibly due to an established systematic collaboration with Wikipedia [20].

Our study shows a ranking of journals according to their actual, practical usability in Wikipedia medical articles, which may be advantageous over journal self-descriptions or preconceptualized categorizations. It allows inclusion of general science and health journals and is an alternative, if not more reliable, measure of journal impact on popular knowledge based on decisions of the self-governed, peer-production community.

Our study has a few limitations. Our analysis was limited to the largest Wikipedia, the English-language one. It is possible that other language Wikipedias might have divergent patterns, as demonstrated by prior research [10]. Where possible, the most widely used unique identifier for scholarly journal articles—the digital object identifier (DOI)—was used [14]. However, some journal articles cited by Wikipedia did not have had an identifier. In such instances, where possible, journals were identified using the publication title, journal name, and Crossref application programming interface. Due to the presence of abbreviations, incomplete entries or identification of misspelling was not

always possible, and therefore, some entries were missed. Books, regarded as secondary sources and thus likely to be preferentially cited on Wikipedia's medical pages as per internal guidelines, were not included in this analysis. In addition, articles published before the year 2000 may have statistically worse coverage by Crossref and the DOI than contemporary articles. There was a bias in our results toward journals that cover clinical topics, because the *Journal of Medical Internet Research* and related journals were cited mainly by Wikipedia pages on eHealth/Health informatics. These pages were predominantly categorized under the fields of computing or technology instead of medicine. To minimize this limitation, future studies could begin with a list of biomedical journals and subsequently search where these studies are cited in Wikipedia, regardless of category. A notable strength of this study was the availability of robust date data for a large number of journal articles. Notably, it is estimated that 694,930 references supporting medical content were published on Wikipedia in 2017, including all kinds of sources [21].

Given the fact that Wikipedia editing increases information literacy [22-24] and that Wikipedia is increasingly adopted by academics [25], we believe that Wikipedia can be relied upon to supplement our knowledge about journal quality. Similar research for other disciplines is warranted.

Acknowledgments

DJ received a research grant (UMO-2012/05/E/HS4/01498) from the Polish National Science Center.

Authors' Contributions

GM conceived the study. DJ, GM, and MW designed the study and its methods. MW performed the analysis. DJ and GM contributed to the literature review. All authors contributed to manuscript writing and interpretation of results, added important intellectual content, and approved the final version. All authors' contributions were equal.

Conflicts of Interest

DJ is a member of The Wikimedia Foundation Board of Trustees. GM is the Assistant to the Editor-in-Chief of the *WikiJournal of Medicine*.

Multimedia Appendix 1

Methods.

[PDF File (Adobe PDF File), 28KB - jmir_v21i1e11429_app1.pdf]

Multimedia Appendix 2

Top 200 journals.

[XLSX File (Microsoft Excel File), 20KB - jmir_v21i1e11429_app2.xlsx]

References

1. Jemielniak D. Common Knowledge? An Ethnography Of Wikipedia. Stanford: Stanford University Press; 2018.
2. Thompson N, Hanley D. MIT Sloan Research Paper. 2017. Science is Shaped by Wikipedia: Evidence From a Randomized Control Trial URL: http://doughanley.com/files/papers/thompson_hanley_wikipedia.pdf [accessed 2018-06-29] [WebCite Cache ID 70X1DFGCh]
3. Tomaszewski R, MacDonald K. A Study of Citations to Wikipedia in Scholarly Publications. Science & Technology Libraries 2016 Aug 02;35(3):246-261. [doi: 10.1080/0194262X.2016.1206052]
4. Jemielniak D, Aibar E. Bridging the gap between wikipedia and academia. J Assn Inf Sci Tec 2016 Apr 04;67(7):1773-1776. [doi: 10.1002/asi.23691]
5. Shafee T, Masukume G, Kipersztok L, Das D, Häggström M, Heilman J. Evolution of Wikipedia's medical content: past, present and future. J Epidemiol Community Health 2017 Nov;71(11):1122-1129 [FREE Full text] [doi: 10.1136/jech-2016-208601] [Medline: 28847845]
6. Herbert VG, Frings A, Rehatschek H, Richard G, Leithner A. Wikipedia--challenges and new horizons in enhancing medical education. BMC Med Educ 2015 Mar 06;15:32 [FREE Full text] [doi: 10.1186/s12909-015-0309-2] [Medline: 25879421]
7. Scaffidi MA, Khan R, Wang C, Keren D, Tsui C, Garg A, et al. Comparison of the Impact of Wikipedia, UpToDate, and a Digital Textbook on Short-Term Knowledge Acquisition Among Medical Students: Randomized Controlled Trial of Three Web-Based Resources. JMIR Med Educ 2017 Oct 31;3(2):e20 [FREE Full text] [doi: 10.2196/mededu.8188] [Medline: 29089291]

8. Reavley N, Mackinnon A, Morgan A, Alvarez-Jimenez M, Hetrick S, Killackey E, et al. Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. Psychol Med 2012 Aug;42(8):1753-1762. [doi: 10.1017/S003329171100287X] [Medline: 22166182]
9. Rajagopalan MS, Khanna VK, Leiter Y, Stott M, Showalter TN, Dicker AP, et al. Patient-oriented cancer information on the internet: a comparison of wikipedia and a professionally maintained database. J Oncol Pract 2011 Sep;7(5):319-323 [FREE Full text] [doi: 10.1200/JOP.2010.000209] [Medline: 22211130]
10. Jemielniak D, Wilamowski M. Cultural diversity of quality of information on Wikipedias. Journal of the Association for Information Science and Technology 2017 Aug 07;68(10):2460-2470. [doi: 10.1002/asi.23901]
11. Holtz P, Fetahu B, Kimmerle J. Effects of Contributor Experience on the Quality of Health-Related Wikipedia Articles. J Med Internet Res 2018 May 10;20(5):e171 [FREE Full text] [doi: 10.2196/jmir.9683] [Medline: 29748161]
12. Lucassen T, Dijkstra R, Schraagen J. First Monday. 2012. Readability of Wikipedia URL: http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3916/3297 [accessed 2018-12-12] [WebCite Cache ID 70X0siZOY]
13. Yasseri T, Kornai A, Kertész J. A practical approach to language complexity: a Wikipedia case study. PLoS One 2012;7(11):e48386 [FREE Full text] [doi: 10.1371/journal.pone.0048386] [Medline: 23189130]
14. Guglielmi G. Wikipedia's top-cited scholarly articles revealed. Nature 2018;557(7705):291. [Medline: 29765126]
15. Masukume G, Kipersztok L, Das D, Shafee TMA, Laurent MR, Heilman JM. Medical journals and Wikipedia: a global health matter. Lancet Glob Health 2016 Nov;4(11):e791 [FREE Full text] [doi: 10.1016/S2214-109X(16)30254-6] [Medline: 27765289]
16. Koppen L, Phillips J, Papageorgiou R. Analysis of reference sources used in drug-related Wikipedia articles. J Med Libr Assoc 2015 Jul;103(3):140-144 [FREE Full text] [doi: 10.3163/1536-5050.103.3.007] [Medline: 26213506]
17. Medicine articles by quality statistics Internet. Wikipedia URL: https://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team/Medicine_articles_by_quality_statistics [accessed 2018-06-28] [WebCite Cache ID 70VyU1XM1]
18. Wikipedia. Recentism URL: https://en.wikipedia.org/wiki/Wikipedia:Recentism [accessed 2018-06-28] [WebCite Cache ID 70VzBf2ax]
19. Wikipedia. Identifying reliable sources (medicine) Internet URL: https://en.wikipedia.org/wiki/Wikipedia:Identifying_reliable_sources_(medicine) [accessed 2018-06-28] [WebCite Cache ID 70VzqO2vd]
20. Mathew ME, Joseph A, Heilman JM, Tharyan P. Cochrane and Wikipedia: the collaborative potential for a quantum leap in the dissemination and uptake of trusted evidence. Cochrane Database Syst Rev 2013 Oct 22(10):ED000069. [doi: 10.1002/14651858.ED000069] [Medline: 24475488]
21. Heilman JM, West AG. Wikipedia and medicine: quantifying readership, editors, and the significance of natural language. J Med Internet Res 2015 Mar 04;17(3):e62 [FREE Full text] [doi: 10.2196/jmir.4069] [Medline: 25739399]
22. Jemielniak D. Wikimedia movement governance: the limits of a-hierarchical organization. Journal of OrgChange Mgmt 2016 May 09;29(3):361-378. [doi: 10.1108/JOCM-07-2013-0138]
23. McKenzie B, Brown J, Casey D, Cooney A, Darcy E, Giblin S, et al. From Poetry to Palmerstown: Using Wikipedia to Teach Critical Skills and Information Literacy in A First-Year Seminar. College Teaching 2018 May 14:1-8 (forthcoming) [FREE Full text] [doi: 10.1080/87567555.2018.1463504]
24. Dawe L, Robinson A. Wikipedia editing and information literacy: a case study. Information and Learning Science 2017 Jan 09;118(1/2):5-16. [doi: 10.1108/ILS-09-2016-0067]
25. Konieczny P. Teaching with Wikipedia in a 21-century classroom: Perceptions of Wikipedia and its educational benefits. J Assn Inf Sci Tec 2016 Mar 15;67(7):1523-1534. [doi: 10.1002/asi.23616]

Abbreviations

COI: conflict of interest
DOI: digital object identifier
JCR: Journal Citation Reports
PNAS: Proceedings of the National Academy of Sciences
SJR: Scientific Journal Ranking

Edited by G Eysenbach; submitted 29.06.18; peer-reviewed by J Heilman, P Konieczny; comments to author 25.07.18; revised version received 03.08.18; accepted 22.08.18; published 18.01.19.

Please cite as:
Jemielniak D, Masukume G, Wilamowski M
The Most Influential Medical Journals According to Wikipedia: Quantitative Analysis
J Med Internet Res 2019;21(1):e11429
URL: http://www.jmir.org/2019/1/e11429/
doi:10.2196/11429
PMID:30664451

Original Paper

Perceptions of Visualizing Physical Activity as a 3D-Printed Object: Formative Study

Sam Graeme Morgan Crossley[1*], BSc, PhD; Melitta Anne McNarry[1*], BSc, PhD; Joanne Hudson[1*], PhD, HCPC; Parisa Eslambolchilar[2*], BEng, MEng, PhD; Zoe Knowles[3*], PhD, HCPC; Kelly Alexandra Mackintosh[1*], BSc, MSc, PhD

[1]Applied Sport Technology Exercise and Medicine Research Centre, School of Sport and Exercise Sciences, Swansea University, Swansea, United Kingdom

[2]Human Factors Technology Research Priority Area, School of Computer Science and Informatics, Cardiff University, Cardiff, United Kingdom

[3]Physical Activity Exchange, School of Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Parisa Eslambolchilar, BEng, MEng, PhD
Human Factors Technology Research Priority Area
School of Computer Science and Informatics
Cardiff University
Queen's Buildings, North Building
5 The Parade, Newport Road
Cardiff, CF24 3AA
United Kingdom
Phone: 44 0 29 2087 9346
Email: EslambolchilarP@cardiff.ac.uk

Abstract

Background: The UK government recommends that children engage in moderate-to-vigorous physical activity for at least 60 min every day. Despite associated physiological and psychosocial benefits of physical activity, many youth fail to meet these guidelines partly due to sedentary screen-based pursuits displacing active behaviors. However, technological advances such as 3D printing have enabled innovative methods of visualizing and conceptualizing physical activity as a tangible output.

Objective: The aim of this study was to elicit children's, adolescents', parents', and teachers' perceptions and understanding of 3D physical activity objects to inform the design of future 3D models of physical activity.

Methods: A total of 28 primary school children (aged 8.4 [SD 0.3] years; 15 boys) and 42 secondary school adolescents (aged 14.4 [SD 0.3] years; 22 boys) participated in semistructured focus groups, with individual interviews conducted with 8 teachers (2 male) and 7 parents (2 male). Questions addressed understanding of the physical activity guidelines, 3D model design, and both motivation for and potential engagement with a 3D physical activity model intervention. Pupils were asked to use Play-Doh to create and describe a model that could represent their physical activity levels (PAL). Data were transcribed verbatim and thematically analyzed, and key emergent themes were represented using pen profiles.

Results: Pupils understood the concept of visualizing physical activity as a 3D object, although adolescents were able to better analyze and critique differences between low and high PAL. Both youths and adults preferred a 3D model representing a week of physical activity data when compared with other temporal representations. Furthermore, all participants highlighted that 3D models could act as a motivational tool to enhance youths' physical activity. From the Play-Doh designs, 2 key themes were identified by pupils, with preferences indicated for models of abstract representations of physical activity or bar charts depicting physical activity, respectively.

Conclusions: These novel findings highlight the potential utility of 3D objects of physical activity as a mechanism to enhance children's and adolescents' understanding of, and motivation to increase, their PAL. This study suggests that 3D printing may offer a unique strategy for promoting physical activity in these groups.

(J Med Internet Res 2019;21(1):e12064) doi:10.2196/12064

KEYWORDS

3D printing; feedback; youth; education; school

Introduction

Background

The UK government recommends that children aged 5 to 18 years engage in at least 60 min of moderate-to-vigorous physical activity (MVPA) every day [1]. Despite the well-established physiological and psychosocial health benefits of regular physical activity for youths [2-5], many fail to meet these recommended guidelines [6]. More specifically, for these populations, sedentary screen-based pursuits are thought to have displaced active behaviors and have been independently associated with adverse health outcomes such as obesity [7] and hypertension [8]. According to Noonan et al [9], there is a lack of understanding within youths on the various forms of physical activity, including those of active travel and unstructured play, with a need to educate how these types of activities contribute to achieving the physical activity recommendations. Conversely, Kremers et al [10] argue that a lack of awareness of physical activity among youths is likely to make them less susceptible to educational programs that are aimed to influence attitudes, norms, self-efficacy, or other cognitive means, as they will not perceive the need to change. Indeed, research supports this notion, demonstrating that youths who are aware of their physical activity levels (PAL) and the recommended guideline are, on average, 20 min more active than their unaware counterparts [10], and as a result, they are more likely to achieve the daily 60 min of MVPA [11,12]. Therefore, developing youths' understanding and awareness of their physical activity behaviors is crucial for implementing a successful health program designed to increase PAL [10].

Schools have been identified as ideal settings to integrate health-promoting interventions because of their established infrastructure and role in health education [13]. Subsequently, researchers have developed numerous school-based interventions that seek to utilize technology as part of the solution, rather than part of the problem [14-18]. Although technology-based interventions have shown promise in improving psychosocial outcomes, efforts to elicit sustainable behavior change have been less consistent [19]. This may, at least in part, be a result of the traditional power structure of the *all-knowing* adult and the *all-learning* child [20], where adults' development of new technology limits the personal opinions of youths when it comes to deciding what technology should be used within a school-based environment [20].

To develop a successful physical activity intervention, an appropriate conceptual health promotion model should be utilized to focus on the most salient characteristics of the target group [21]. One such model, which is specifically relevant to children's physical activity, is the Youth Physical Activity Promotion Model (YPAPM) [22]. This model provides a comprehensive and structured assessment of the target population's own needs and barriers to participation in physical activity, acknowledging children as the experts [23], and allowing intervention design through the eyes of the child rather than the researcher, teacher, or parent [9]. As argued by Druin [20], children as design partners can play an impactful role in the creation of new technologies that are not only going to be effective and meaningful but also will excite children and aid learning.

Research shows that 80% of youths are visual and tactile learners [24]; therefore, relying simply on numbers and figures as a source of knowledge is limited [25], and richer ways of data representation are required [26]. Indeed, visualizations can play a key role in motivating individuals to enhance their PAL, enabling reflection on personal performance and current level of physical activity [27]. A recent school-based intervention using glanceable light-emitting diode (LED) technology to display groups' PAL reported that children wanted more personalized forms of visual feedback [18], with others suggesting that material rewards are cherished more than virtual rewards [28] because of their higher visibility and uniqueness [29,30]. Indeed, previous research utilizing paper and LED lights to create PA awareness promoting artifacts found that youth took incremental steps toward self-regulation through goal setting and reflection [31]. It was concluded that although the artifacts did not elicit improved physical activity in youth, using tangible artifacts in conjunction with wearables could benefit youths' health [31]. However, it could be argued that paper artifacts do not provide youth with an adequate haptic and proprioceptive experience of personalized feedback to reap health benefits [32]. With the recent rise of the *maker movement* and cost-effective 3D printers [33], numerous opportunities in health-related research have emerged, utilizing 3D printers to create tangible visualizations of physical activity [34-36]. As Jansen et al [37] advocate, there are many benefits of tangible visualizations over on-screen visualizations of data, which include (1) allowing for a more active perception, (2) leveraging nonvisual senses such as touch, (3) integration with the physical world, and (4) harnessing the interplay between vision and touch to facilitate cognition. For example, Khot et al [38] transformed adults' heart rate data into 3D-printed artifacts, with participants reporting that the artifacts acted as a reward and allowed reflection and reminiscence on past physical activities [38]. Indeed, tangible interfaces have been reported to involve children in playful learning [39], engagement, and reflection [40]. Consistent with goal-setting theory [41], incentives are important in maintaining interest in an activity, with incentive-based interventions to *nudge* healthy behavior change in youths demonstrating potential [42,43]. However, whether personalized 3D-printed objects can be used to enhance youths' understanding, awareness, and motivation relating to engagement in physical activity remains to be elucidated.

Aims

Therefore, the aims of the present study were to (1) formatively elicit children's, adolescents', teachers', and parents' perceptions of physical activity data when represented as 3D-printed objects; (2) ascertain how youths visualize their personal 3D objects of physical activity using Play-Doh; (3) obtain parents' and teachers' views on the perceived benefits and barriers of

3D-printed objects of physical activity for youths; and (4) use these data to subsequently inform the design of 3D models and a school-based physical activity intervention.

Methods

Recruitment

In total, 20 primary and secondary schools from the Swansea region of South Wales were contacted and invited to take part. The schools were stratified into high and low socioeconomic status (SES) according to the percentage of students per school eligible to receive free school meals [44]. From those schools that expressed an interest (35%, 7/20) response rate), 4 schools, 1 high- and 1 low-SES primary and secondary schools, were selected based on order of availability to take part in the study. Overall, 27 primary school children (aged 8.4 [SD 0.3] years; 15 boys) and 42 secondary school adolescents (aged 14.4 [SD 0.3]; 22 boys), 8 teachers (2 male), and 7 parents (2 male) provided written informed parental or carer consent and child assent, as appropriate, to participate in the study. All procedures were approved by the Swansea University A-STEM Ethics Committee and were conducted in accordance with the Declaration of Helsinki (reference number: PG/2014/40).

Procedures

All semistructured focus group discussions and interviews were conducted by the first author (SGMC) in a nondirective and unbiased way [45], with 6 groups of children, 8 groups of adolescents, and a total of 13 individual interviews with teachers and parents. Sample questions for the focus groups and one-to-one interviews are presented in Table 1. On 2 separate occasions, 2 parents and 2 teachers were interviewed together because of restricted availability [46]. Focus group discussions with youths involved 4 to 6 participants to allow for lively, yet manageable, interactions [45,47,48], with the exception of 1 primary school focus group where a child with special educational needs required a smaller group of 3 children with 1 support teacher. Both single- and mixed-sex focus groups were conducted [49]. All focus group sessions were completed within the school environment, either within a familiar classroom or in the school library, to provide comfort and reduce anxiety [50]. Participants were seated in a circular arrangement around a table to create a relaxed and informal atmosphere [45], maximizing social interaction and observer involvement [51]. Moreover, this seating arrangement allows the facilitator to sit among the participants to establish a nonauthoritarian approach to questioning. To ensure each of the group members was comfortable with talking aloud and to create an environment in which sharing and listening were valued, an icebreaker question was used [52]. The semistructured focus group questions were informed by enabling, reinforcing, and predisposing factors from the YPAPM [22] to explore physical activity engagement and identify any barriers toward 3D-printed objects in an age-appropriate manner. All predetermined questions were reviewed and discussed by SGMC, MAM, PE, and KAM, and additional feedback was provided independently by 2 Health and Care Professions Council–registered practitioner psychologists (JH and ZK).

Alongside focus group discussions and one-to-one interviews, children, adolescents, and adults were all shown a custom-made video on the concept of 3D printing physical activity. Following this, participants were shown 3 different prototype 3D-printed models displaying example accelerometry-derived physical activity data, and discussions focused on how participants thought the physical activity data were represented by these models. Finally, children and adolescents were asked to independently design their own personalized model of physical activity using Play-Doh. The Play-Doh modeling process builds on the principles of the write, draw, show and tell method [9] by replacing the write and draw components of the framework with the modeling of Play-Doh. Following the Play-Doh modeling task, the facilitator asked each child to articulate and explain the characteristics of their design in a verbal statement at their own pace. All Play-Doh models were photographed for further analyses.

Focus group discussions lasted between 60 and 90 min and 50 and 60 min for primary and secondary school groups, respectively, and adult interviews lasted approximately 25 to 45 min. All the focus groups and one-to-one interviews were digitally voice-recorded (Olympus DM-520 Digital Voice Recorder; Shinjuku, Japan) and video-recorded (Sony Handycam HDR-PJ540, Minato, Japan).

Data Analysis

All focus group discussions and one-to-one interviews were transcribed verbatim, resulting in 774 pages (327, 297, and 150 pages for children, adolescents, and adults, respectively) of raw data. Researchers SGMC, MAM, and KAM read each transcript to familiarize themselves with the data. Transcripts were thematically analyzed by SGMC using data coding and identification of themes [53]. Transcripts were first deductively analyzed using aspects of the YPAPM as a thematic framework [22].

Table 1. Example focus group and interview questions.

Interview	Topic	Example
Children and adolescents	Motivation	What would you think if I said we could 3D print your own personal model, which shows how physically active you are?
Children and adolescents	Model design	What sort of model would you like to develop or represent your own physical activity as in the video, how would it look?
Adults	Motivation	How do you think the 3D models of physical activity could motivate children to be more physically active?
Adults	Model design	Are there any models that you think would be good to help children to visualize physical activity?

Additional emergent themes were then further explored using an inductive process. Both deductive and inductive processes used a manual cut-and-paste technique to identify key themes. Participants' verbatim quotations were chosen by SGMC and discussed in collaboration with MAM and KAM. A frequency count for the meaningful quotes was conducted to record how many participants responded within emergent themes. The themes, meaningful quotations, and frequency counts were then displayed diagrammatically using a pen profile approach. Pen profiling has been used within studies exploring perceptions and experiences of physical activity in youths [47,54] and is considered to be an accessible technique for researchers who have both quantitative and qualitative backgrounds [55]. Through the process of reverse triangulation, authors critically questioned and cross-examined the data in reverse from the pen profiles to the transcripts. This process was repeated, allowing authors to offer alternative interpretations of the data, until a consensus was reached to finalize the pen profile designs. In some cases, visual illustrations were presented to add more context to the data collected. Triangulation of the data tests the robustness of the findings and ensures methodological rigor using a *trustworthiness criterion* [56]. The criterion places trust in the researcher responsible for data collection to determine key findings that are worthy of attention. These were then assessed by PE, JH, and ZK who were not as directly involved in the analysis process [57]. In addition, the primary and secondary school participants' Play-Doh model photos aligned with the relevant verbal statements were analyzed by SGMC, MM, JH, PE, and KAM as a group to identify common trends and designs. Specifically, all Play-Doh model photos, with their respective verbal statements, were displayed on a large white board and appraised by the research team. Throughout this process, the Play-Doh models were grouped based on similar structural (eg, sun or bar chart design) and verbal (eg, the more physical activity you do, the larger the model) characteristics. The most common Play-Doh model designs created by children (abstract, 12/28; graphical, 15/28) and adolescents (graphical, 28/42) were subsequently considered for further interpretation and 3D model design.

Results

Perceptions and Designs of 3D Physical Activity Models

In total, 3 separate pen profiles were constructed to represent children's (Figure 1), adolescents' (Figure 2), and combined parents' and teachers' (Figure 3) perceptions of 3D models. There were consistent themes identified between parents and teachers, and therefore, their data were combined for final analysis.

Children's Perceptions and Designs of 3D Physical Activity Models

As shown in Figure 1, key emergent themes were structured around "Temporal Representation of Physical Activity," "Motivation," "Interpretation," and "Physical Activity Guidelines." The higher order theme "Interpretation" was linked to further subthemes "Physical Activity Representation" and "Design." Primary school children demonstrated the ability to interpret and apply the different component lengths and sizes of the prototype 3D models in relation to physical activity parameters. Specifically, 92% (25/28) of children were able to accurately understand how the changing length of the model represented increasing levels of physical activity. However, only 26% (7/28) of children were able to understand the alternative method of increasing the size of the model to represent greater levels of physical activity. The physical activity data displayed on the models were correctly identified by 59% (16/28) of the children as representing either hours or days of physical activity. The majority of children (81%, 22/28) preferred the 3D models to represent a week of their physical activity data, compared with a day (3/28, 11%), year (2/28, 7%), or month (1/28, 4%):

> *Because you do...you probably do more exercise in a week than a day.* [G16]

From the Play-Doh modeling task, 2 subthemes emerged, one being *abstract* and the other *graphical*. Children revealed no preference for abstract (12/28, 44%) or graphical (15/28, 56%) model representations of physical activity. Children's abstract models were characterized by the model changing shape or size, such as a volcano with more lava erupting for higher levels of physical activity (Figure 4). Graphical representations, such as the flower (Figure 5), distinguished between different hours, days, or weeks of physical activity completed (ie, the flower's petals resembling the different days of physical activity).

A total of 21 children (78%, 21/28) commented that the 3D models had potential to motivate themselves to engage in more physical activity, substantially outweighing the negatives expressed by 1 child. Specifically, children revealed that the 3D models would add competition between classmates and motivate them to do more. For example:

> *Because you might see how people have done much more [physical activity] than you and then you...you would think I want to be like that person and then you'd do more.* [G16]

Overall, 16 children (59%) displayed limited knowledge of the current UK government physical activity guidelines or how to achieve them, with only 3 children able to express the amount through the context of time spent being physically active, with no reference to intensity level. For example:

> *...probably something like an hour, two hours a day...* [PL9]

Figure 1. Children's pen profile. B: boy; G: girl; PA: physical activity; n: frequency counts.

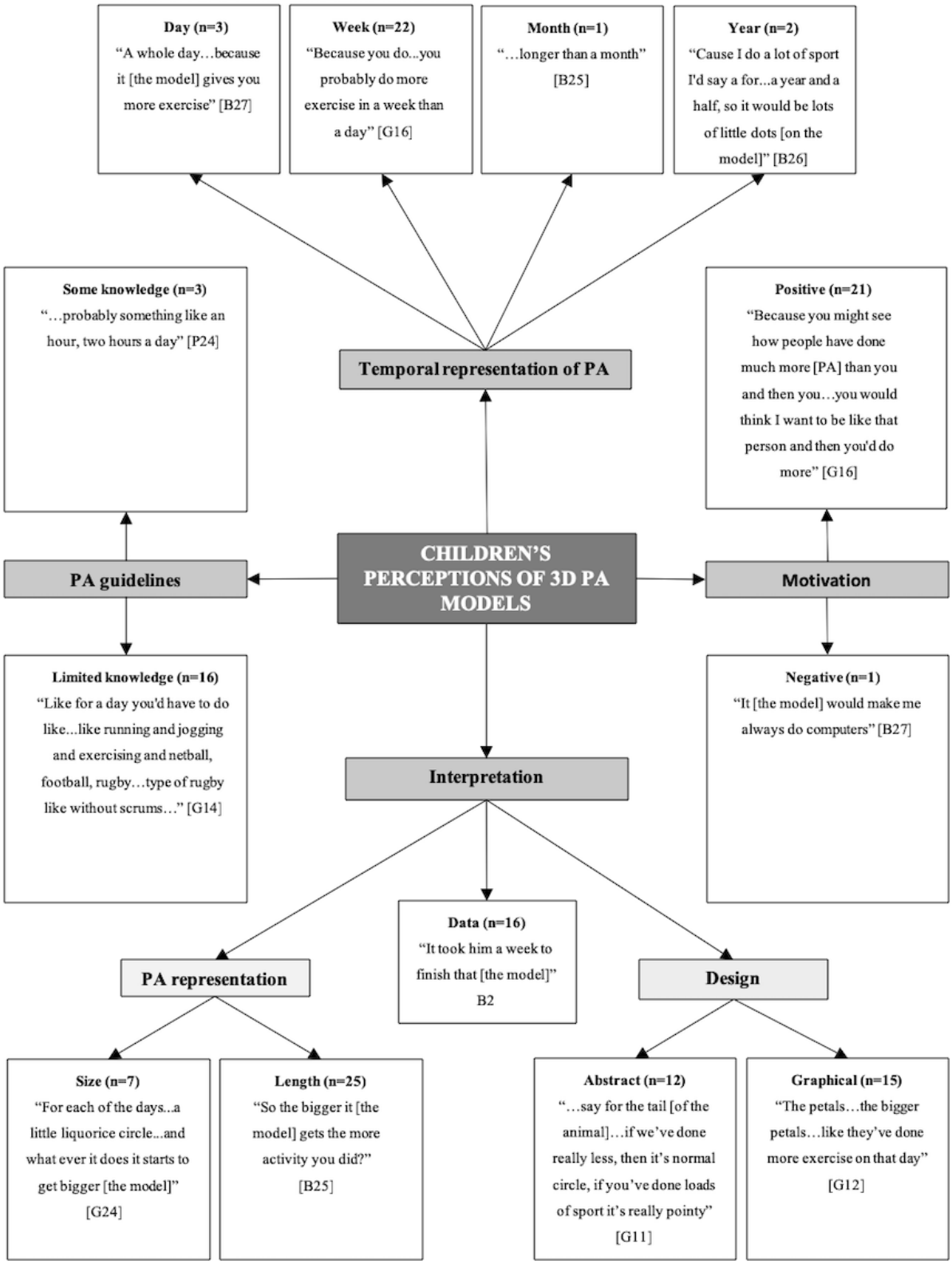


Figure 2. Adolescents' pen profile. B: boy; G: girl; PA: physical activity; n: frequency counts.

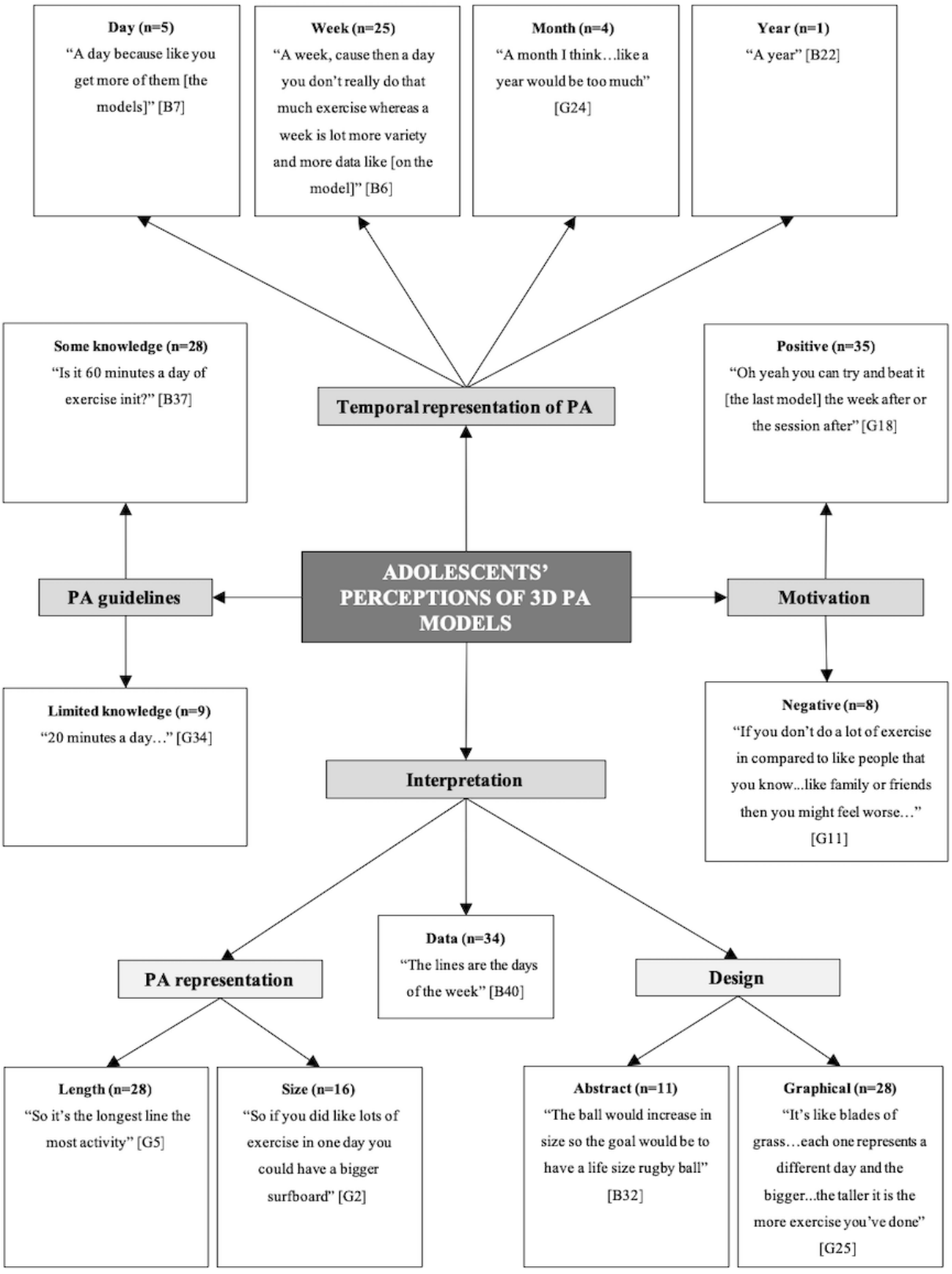


Figure 3. Adults' pen profile. T: teacher; P: parent; M: male; F: female; PA: physical activity; n: frequency counts.

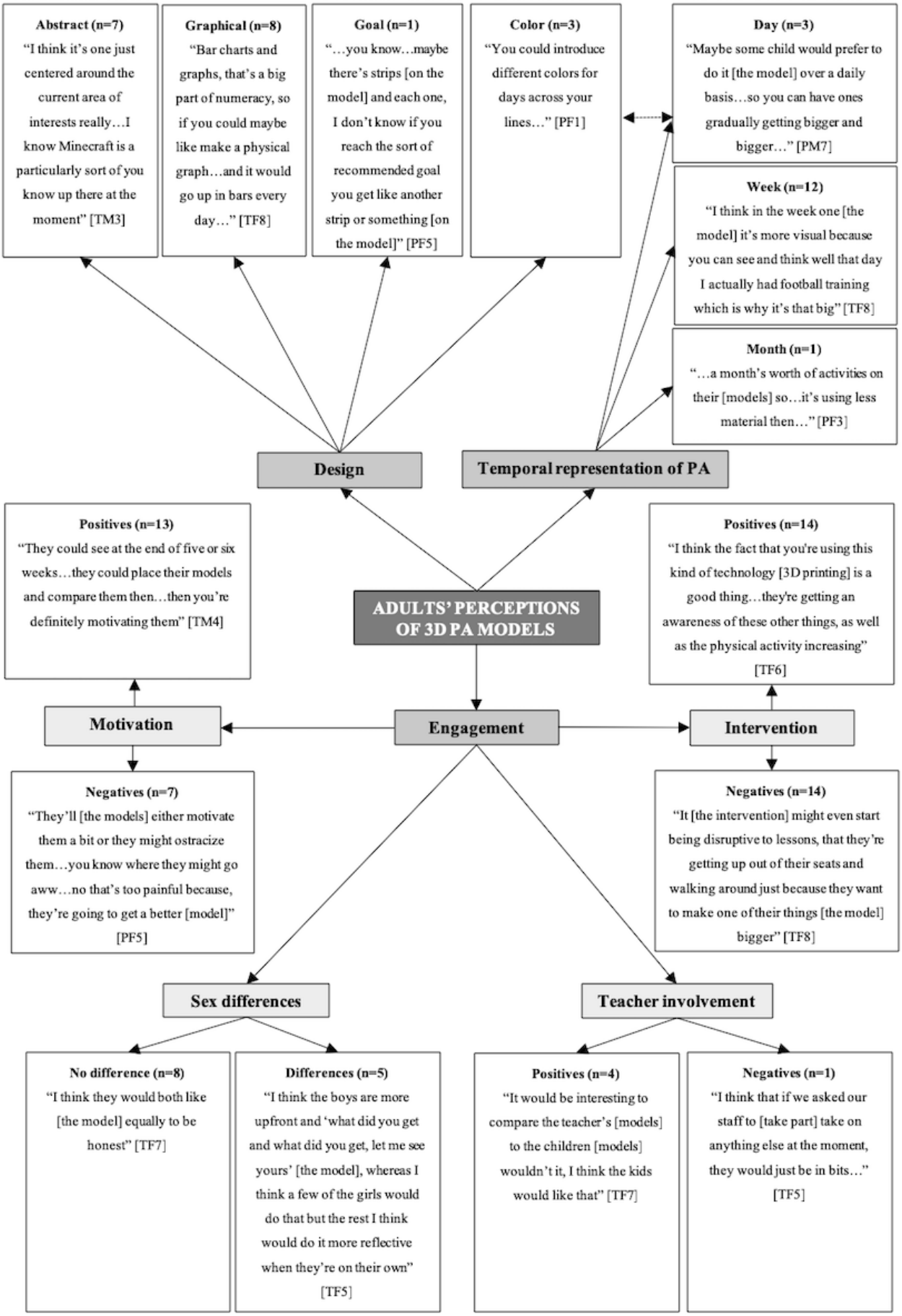


Figure 4. Child's abstract volcano model design.

Figure 5. Child's graphical flower model design.

Adolescents' Perceptions and Designs of 3D Physical Activity Models

Overall, 4 higher order themes were identified structured around "Temporal representation," "Motivation," "Interpretation," and "Physical activity guidelines" (Figure 2). The higher order theme *Interpretation* was further linked to subthemes *Physical activity representation* and *design*. Adolescents demonstrated the ability to identify and compare the different components of the prototype 3D models and their changing length and size in relation to physical activity. Specifically, the increasing size (16/42, 38%) and length (28/42, 67%) of the models were correctly interpreted as representing higher PAL. The majority (34/42, 81%) of adolescents showed a clear understanding of the represented data on the models. For example:

> *The lines [on the models] are the days of the week.* [B30]

> *...so does that mean he's most active Tuesday, Wednesday, Thursday sort of thing.* [G3]

Adolescents highlighted a preference for a week (25/42, 60%) of physical activity data to be displayed on the 3D models because of the greater variety and reflection of their PAL in a week when compared with a model based on a day (5/42, 12%), month (4/42, 10%), or year (1/42, 2%). The Play-Doh modeling task displayed similar subthemes to those found in children, with a larger proportion of designs displaying graphical (28/42, 67%) compared with abstract (11/42, 26%) designs. Abstract models, such as the butterfly (Figure 6), were characterized by the changing size or detail of the models. Graphical representations resembled typical bar charts or line graphs (Figure 7) to display different days, weeks, or months of physical activity.

A total of 35 adolescents (83%, 35/42) expressed that the 3D models would motivate them to engage in more physical activity by beating previous models. For example:

> *Oh yeah you can try and beat it [the model] the week after or the session after.* [G18]

Overall, 8 adolescents (19%, 8/42) thought that the 3D models may discourage engagement in physical activity because of feelings of doing worse than others and embarrassment if the model showed low PAL:

> *If you don't do like a lot of exercise in compared to like people that you know...like family or friends then you might feel worse...* [G11]

> *...if other people like saw the object or something it might be a bit embarrassed if you haven't done enough exercise.* [G21]

In total, 28 adolescents (67%, 28/42) showed some knowledge of the government guidelines for physical activity. A specific Sport Wales initiative called *5x60* [58] may have influenced these findings:

> *...they [Sport Wales officers] try and get everyone to do five sessions of sixty minutes a week of exercise.* [G3]

> *...yeah what's it called...five sixty...five hours of sixty...no five lots of sixty minutes per week...* [G18]

Figure 6. Adolescent's abstract butterfly model design.

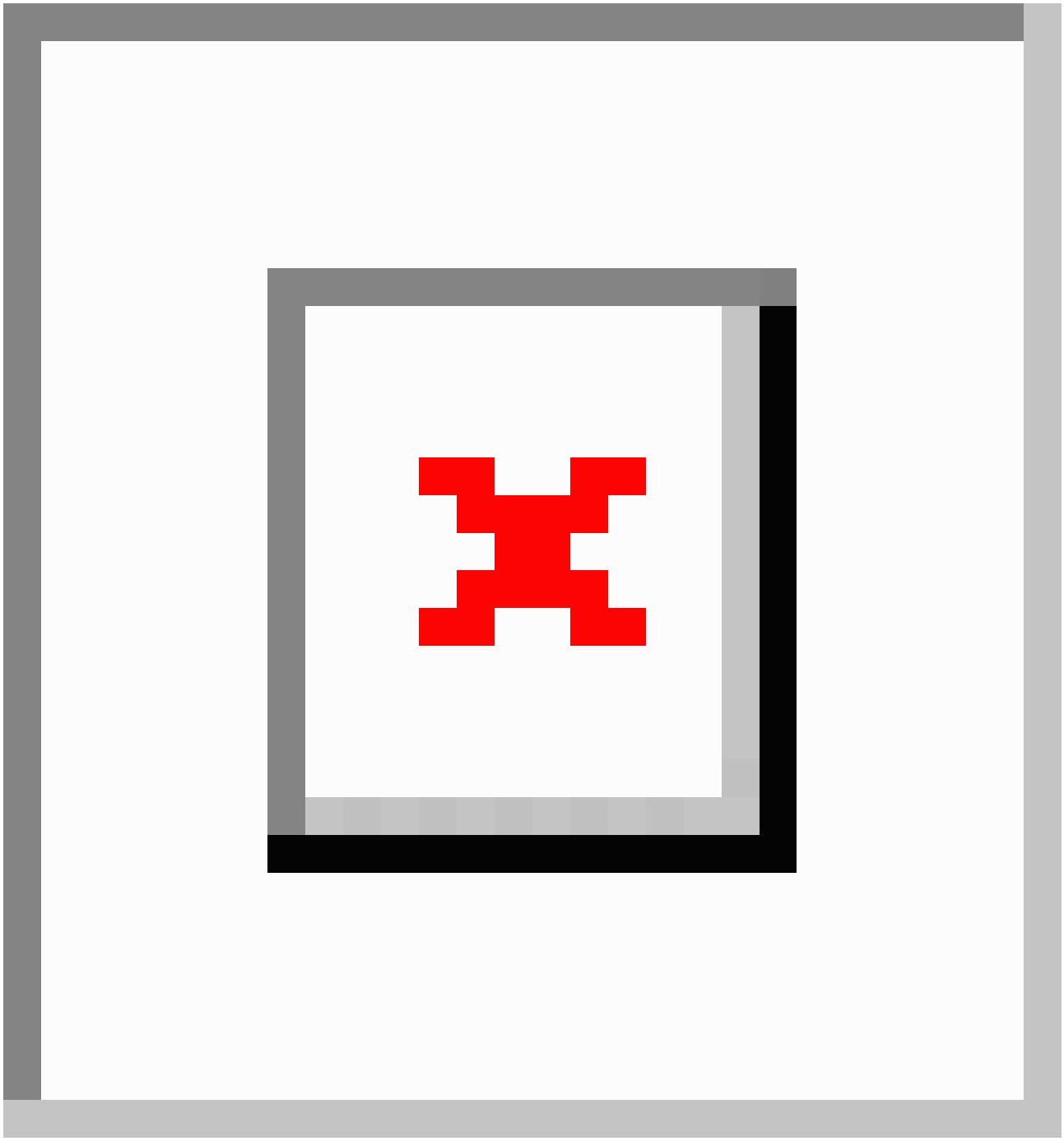
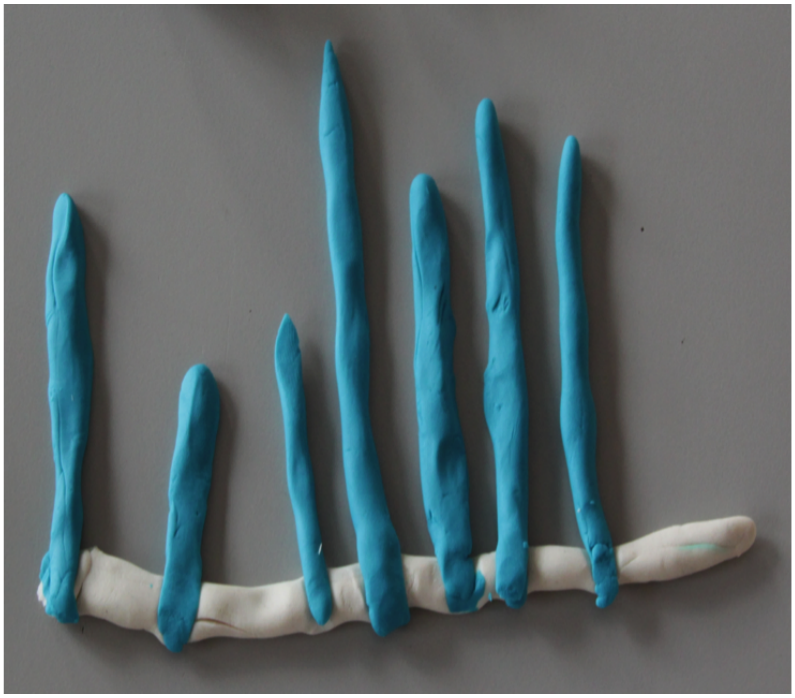


Figure 7. Adolescent's graphical bar chart model design.



Adults' Perceptions and Designs of 3D Physical Activity Models

The key adult emergent themes were "Design," "Temporal Representation of PA," "Engagement," and "Motivation," with a few distinct subthemes (Figure 3). The higher order theme *Engagement* was linked to additional subthemes *Sex differences*, *Teacher involvement*, and *Intervention*. Adults described characteristics similar to those used to construct both abstract (7/15, 47%) and graphical (8/15, 53%) model representations. Adults placed emphasis on making the 3D models attractive and recognizable but also something that challenges children's and adolescents' numeracy skills to work within the school curriculum:

> *Bar charts and graphs, that's a big part of numeracy, so if you could maybe like make a physical graph...and it would go up in bars every day* [TF8]

Moreover, 1 parent added that a link between the 3D model and a recommended goal for physical activity could help encourage youths to achieve greater PAL:

> *...you know...maybe there's strips [on the model] and each one, I don't know if you reach the sort of recommended goal you get like another strip or something [on the model]*. [PF5]

Similar to youths, adults preferred a week (12/15, 80%) of physical activity data represented on the model, as this was thought to have greater potential to visually guide youths, creating more awareness of their physical activity behaviors than a day (3/15, 20%) or month (1/15, 7%). Furthermore, some adults emphasized that changing the color (3/15, 20%) of lines on the models could visually aid participants in distinguishing the different days. The majority (13/15, 87%) of adults believed that if youths received and compared new models over time, this would act as a strong motivation for increased engagement in physical activity:

> *They [youths] could see at the end of five or six weeks...they could place their models and compare them thenthen you're definitely motivating them [youths]*. [TM4]

Furthermore, some teachers (4/8, 50%) reported that receiving their own 3D models would act as an additional competition and potential motivation for the pupils. However, some adults (7/15, 47%) expressed that the 3D models may ostracize youths from others if they underachieved in physical activity.

> *they [the models] might ostracize them...you know where they might go...no that's too painful because, they're going to get a better [model]*. [PF5]

Adults perceived both positives (14/15, 93%) and negatives (14/15, 93%) for participants' engagement with the concept of 3D printing physical activity. Positives included that the use of new technology (ie, 3D printers) would create awareness of current technological advances, with negative responses highlighting concerns about potential disruptions to teaching during lesson time. Moreover, some adults (8/15, 53%) believed that there would be no differences in how boys and girls engaged with the models, although 5 adults (33%, 5/15) highlighted that the girls may be more reflective and the boys more competitive:

> *I think the boys are more upfront and "what did you get and what did you get, let me see yours [the model]", whereas I think a few of the girls would do that but the rest I think would do it more reflective when they're on their own.* [TF5]

Discussion

Principal Findings

The aims of the research were first, to formatively elicit children's, adolescents', teachers', and parents' perceptions of physical activity data when represented as 3D-printed objects and their personal designs and to examine the perceived benefits and barriers to intervention participation. This research extends from that of previous studies that have implemented formative research techniques to inform the development of school-based interventions [47,59]. The second aim of the study was to use the formative data to inform the design of 3D models of physical activity to enhance youths' understanding, awareness, and PAL.

The data indicated that youths can conceptualize physical activity data represented as a 3D object. This ability to detect and mentally represent a relationship between a symbol (ie, 3D object) and its referent (ie, physical activity) is known as representational insight [60]. However, the visual nature of the models does not always guarantee representational insight and its relation to the intended use [60]. For example, adolescents in this study showed greater ability to analyze and critique the physical activity behaviors represented on the prototype models. Adolescents could highlight, in some detail, differences in low and high PAL and how this related to their own and others' personal habits. These differences between adolescents and children could be explained by a greater age-related cognitive ability in adolescents [61]. However, differences in cognitive ability may be less influential, as evidence suggests that visualizations help make complex information more accessible and cognitively tractable [60]. More specifically, previous research supports the use of tangible objects to stimulate youths' intellectual development as they support a more natural way of learning [32,39,40,62,63], aligning with youths being regarded as *visual and tactile* learners [24]. For example, Gillet et al [32] investigated the use of 3D-printed enzyme molecules for teaching biology in youth, reporting that the tangible models provided a natural and intuitive mechanism for manipulation, exploration, and a proprioceptive pathway for learning. Although these findings hold promise, given that youths recognize the relationship between the tangible visualization and its intended referent, which is a necessary condition for developing a visual learning tool, others argue that an isolated approach is not sufficient [60,64-66]; it is important that youths understand the meaning and importance of the concepts represented on the visualizations to enable increased awareness of their personal physical activity behaviors [60]. In this light, future research should consider investigating 3D-printed physical activity feedback conditions to include and exclude an additional classroom educational component on PAL to fully understand the benefits of the 3D model alone.

This study revealed that youths believed the 3D models would act as a motivational tool to enhance their own PAL and that of their peers. Indeed, previous research suggests that school-based interventions that promote youths' physical activity with the presence of peers significantly increase their motivation for physical activity [67] as well as their enjoyment [68,69], intensity [70], and engagement in out-of-school physical activity [71]. Furthermore, the majority of primary school children expressed that the 3D models would introduce competition between classmates, motivating them to engage in more physical activity. It has been argued that competition between children can be healthy if it provides feedback about performance and improvements, where children can learn about themselves, and the sole or primary objective is not about winning [72]. Conversely, adolescents placed more emphasis on how they would be motivated by beating their own personal model from the week before rather than comparing with others. These differences between youths could be, in part, explained by the adolescents' greater understanding of the concept of effort in the physical domain [73] and applied ability to think independently, fostering enhanced self-evaluation skills that are important for preparation into adulthood [74]. Parents and teachers also agreed that the models would help motivate children and adolescents, allowing them to compare the models over time. Adults highlighted that boys may take a more competitive approach than girls who may engage in more reflective thoughts about the 3D models. Indeed, evidence suggests that young males engage in more individualistic competition than female counterparts [75]. Contrary to this, Bjorkqvist [76] found that girls use subtler, more indirect strategies for competition than boys from childhood to adulthood. Adolescents also displayed concerns that they might be perceived as inactive by their significant others, a consensus that was supported by the adults. Similar concerns have been raised when using digital fish avatars, the growth and emotional state of which is dependent on the participant's PAL, with participants reporting being discouraged from using the app if they saw that the fish avatar was unhappy [77]. Therefore, monitoring how youths personally evaluate models displaying low PAL, and their support and interactions with significant others should be considered further. Beyond the scope of this study, it is pertinent to note that further research is also required to adapt these models to other populations and cultures, with the current results suggesting that children with special educational needs may misinterpret the models with negative health consequences for PAL, such as increased engagement with computer-based behaviors.

For the adults, the tangibility of being able to hold something that participants have created was perceived as original and personalized. Adults expressed that the tactile forms of information would interest youths and encourage them to purposely think about the importance of physical activity, as previously identified by Mackintosh et al [18]. Furthermore, they also believed that the 3D models could act as a material reward or medal representing achieved physical activity:

> *[something children and adolescents] could put [the models] up on their wall when they get them.* [PM7]

Indeed, much research suggests that material rewards are cherished more than virtual rewards [28], as a result of their higher visibility and low replication possibility [29,30]. Incentive-based interventions to encourage youth to take part in more physical activities have been shown to have promising effects [43,78], although findings have been mixed regarding sustained behavior change following removal of incentives [79]. Sport capitalizes on this incentive form of reward system with

physical medals and trophies being presented to individuals. However, although these rewards focus on the completion of certain fitness or sports goals, they do not embody any personal data or represent the active self [80]. Khot et al [80] do, however, note that there is a learning value to be gained from *blending* rewards and representations to create more personalized and meaningful data. This concept is supported by findings from *Pokémon GO*, where children and adolescents can create and identify themselves with a visual avatar surrounded by recognizable characters (eg, Pikachu) in a socially networked system, which was associated with significant increases in physical activity in both age groups [81].

The current utilization of Play-Doh enabled youths to creatively explore, adapt, and develop their personal 3D model creations. This relatively inexpensive form of design prototyping has been used previously with malleable materials and is effective for brainstorming new ideas and designs from which high-tech prototypes emerge [20]. Our study's findings revealed that children and adolescents preferred different types of 3D model design, leading to the development of 2 age-specific 3D models of physical activity. For children, a preference for a combination of both abstract (43%, 12/27) and graphical (54%, 15/27) models was demonstrated, most commonly expressed as Play-Doh models of flower- or sun-like shapes. However, to avoid any potential sex bias resulting in boys dissociating with a flower-shaped 3D model, the more neutral sun-shaped 3D model design was chosen for further development. Interestingly, a majority of adolescents (67%, 28/42) showed a preference through Play-Doh models for a simple bar chart design. However, with regard to the 2 different age-specific 3D models identified, there is limited literature as to whether the mapping of data should be abstract or graphical. Abstract data allow users to be more curious and speculative, whereas graphical representations provide more direct and comprehensive representations of data. Davis et al [82] suggest that more informative feedback provides greater opportunities to learn and improve performance. Indeed, it has been shown that tangible bar charts have benefits for information recall when compared with digital visualizations [83]. Contrary to this, more abstract methods of feedback may provide more positive engagement and support [84]. Anderson et al [85] also suggest that abstract visualizations increased motivation to achieve higher PAL in adults. Subsequently, adults believed that both methods of mapping a week of physical activity data were equally important, adding that presenting daily physical activity could potentially *overwhelm* the children and adolescents with data. As Khot et al [80] pointed out, embedding too much data can make the material model less readable, but on the other hand, with too little data, the model loses its intended purpose.

Although physical activity recommendations for youths are set to advise them on how to achieve an active lifestyle and create awareness of the important health benefits, few children were able to identify the UK-recommended amount of physical activity. Children's interpretations of how much physical activity they should achieve were largely based on *how much sport* or *how many different sports* they could complete per day (eg, football, rugby, netball, and running), aligning with previous research findings [86]. In comparison, the adolescent group showed greater knowledge of the government guidelines, but this may have been influenced by the ongoing Sport Wales initiative 5x60 [58] implemented at the time of the study and aimed at encouraging youths to engage in 60 min of MVPA every day within school. However, it was evident that neither children nor adolescents were able to associate their understanding of the UK government recommendations with the intensity levels of MVPA, which highlights the need to promote youths' knowledge of government recommendations, as reported by Mackintosh et al [18]. As aforementioned, tangible interfaces may offer a more playful learning experience [39] and natural interaction than other learning interfaces [87-89], suggesting that the tangibility of data may benefit children's and adolescents' learning [62]. As 1 parent expressed, creating a recommended goal for the youths on the model could be beneficial. Therefore, using a goal-setting strategy [41] and structurally developing the government recommendation into a tangible goal on the model may not only help in developing children's and adolescents' understanding of the government recommendations of 60 min of MVPA but also motivate youths to increase their PAL.

Limitations

One of the major strengths of this study is its originality; however, this also highlights the paucity of other supporting research for this age group and that further investigation is warranted on this tangible form of data representation. Research should focus on the relative effectiveness of different types of 3D-printed visualizations of physical activity for the promotion of active learning in youths and as a means of strengthening the articulation of such initiatives with public health guidelines (ie, 60 min of MVPA) to enhance understanding and increase the motivation and engagement of youths in sustained physical activity.

Conclusions

This formative study provides insight into the utilization of tangible 3D-printed objects displaying physical activity as a tool to benefit children and adolescents. The findings demonstrate how youths actively and enthusiastically engaged with the concept of 3D objects of physical activity and felt it could not only enhance their understanding of, but motivate them to increase, their PAL. From pupils' Play-Doh model outputs, 2 age-specific 3D models representing weekly physical activity data were developed. The results of the formative research support the design of school-based physical activity interventions that utilize 3D printing of youths' personal data as a unique strategy to promote their engagement in physical activity.

Acknowledgments

The authors thank all the teachers and parents for their help and the children who participated in the study.

Conflicts of Interest

None declared.

References

1. Department of Health and Social Care. London; 2011 Jul 11. Start active, stay active: report on physical activity in the UK URL: https://www.gov.uk/government/publications/start-active-stay-active-a-report-on-physical-activity-from-the-four-home-countries-chief-medical-officers [WebCite Cache ID 73BamUHJY]
2. Sothern MS, Loftin M, Suskind RM, Udall JN, Blecker U. The health benefits of physical activity in children and adolescents: implications for chronic disease prevention. Eur J Pediatr 1999 Apr;158(4):271-274. [Medline: 10206121]
3. Nieman P. Psychosocial aspects of physical activity. Paediatr Child Health 2002 May;7(5):309-312 [FREE Full text] [Medline: 20046307]
4. Janssen I, Leblanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010 May 11;7:40 [FREE Full text] [doi: 10.1186/1479-5868-7-40] [Medline: 20459784]
5. Eime RM, Young JA, Harvey JT, Charity MJ, Payne WR. A systematic review of the psychological and social benefits of participation in sport for children and adolescents: informing development of a conceptual model of health through sport. Int J Behav Nutr Phys Act 2013 Aug 15;10:98 [FREE Full text] [doi: 10.1186/1479-5868-10-98] [Medline: 23945179]
6. Hills AP, Andersen LB, Byrne NM. Physical activity and obesity in children. Br J Sports Med 2011 Sep;45(11):866-870. [doi: 10.1136/bjsports-2011-090199] [Medline: 21836171]
7. Proctor MH, Moore LL, Gao D, Cupples LA, Bradlee ML, Hood MY, et al. Television viewing and change in body fat from preschool to early adolescence: the Framingham Children's Study. Int J Obes Relat Metab Disord 2003 Jul;27(7):827-833. [doi: 10.1038/sj.ijo.0802294] [Medline: 12821969]
8. Pardee PE, Norman GJ, Lustig RH, Preud'homme D, Schwimmer JB. Television viewing and hypertension in obese children. Am J Prev Med 2007 Dec;33(6):439-443. [doi: 10.1016/j.amepre.2007.07.036] [Medline: 18022058]
9. Noonan R, Boddy LM, Fairclough SJ, Knowles ZR. Write, draw, show, and tell: a child-centred dual methodology to explore perceptions of out-of-school physical activity. BMC Public Health 2016 Apr 14;16:326 [FREE Full text] [doi: 10.1186/s12889-016-3005-1] [Medline: 27080384]
10. Kremers SP, Dijkman MA, de Meij JS, Jurg ME, Brug J. Awareness and habit: important factors in physical activity in children. Health Education 2008;108(6):475-488. [doi: 10.1108/09654280810910881]
11. Roth M, Stamatakis E. Linking young people's knowledge of public health guidelines to physical activity levels in England. Pediatr Exerc Sci 2010 Aug;22(3):467-476. [Medline: 20814041]
12. Nemet D, Geva D, Meckel Y, Eliakim A. Health-related knowledge and preferences in low socio-economic kindergarteners. Int J Behav Nutr Phys Act 2012 Jan 10;9:1 [FREE Full text] [doi: 10.1186/1479-5868-9-1] [Medline: 22233712]
13. Dobbins M, De Corby K, Robeson P, Husson H, Tirilis D. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6-18. Cochrane Database Syst Rev 2009 Jan 21(1):CD007651. [doi: 10.1002/14651858.CD007651] [Medline: 19160341]
14. McDougall J, Ducan M. Children, video games and physical activity: an exploratory study. Int J Disabil Hum Dev 2008;7(1):89-94. [doi: 10.1515/IJDHD.2008.7.1.89]
15. Duncan M, Staples V. The impact of a school-based active video game play intervention on children's physical activity during recess. Hum Mov 2010;11(1):95-99. [doi: 10.2478/v10038-009-0023-1]
16. Poole ES, Miller AD, Xu Y, Eiriksdottir E, Catrambone R, Mynatt ED. The place for ubiquitous computing in schools: lessons learned from a school-based intervention for youth physical activity, in Proceedings of the 13th international conference on Ubiquitous computing. In: Proceedings of the 13th international conference on Ubiquitous computing. 2011 Presented at: UbiComp '11; September 17 - 21, 2011; Beijing, China p. 395-404. [doi: 10.1145/2030112.2030165]
17. Lubans D, Smith JJ, Peralta LR, Plotnikoff RC, Okely AD, Salmon J, et al. A school-based intervention incorporating smartphone technology to improve health-related fitness among adolescents: rationale and study protocol for the NEAT and ATLAS 2.0 cluster randomised controlled trial and dissemination study. BMJ Open 2016 Dec 27;6(6):e010448 [FREE Full text] [doi: 10.1136/bmjopen-2015-010448] [Medline: 27354070]
18. Mackintosh K, Niezen G, Eslambolchilar P. Mission possible: using ubiquitous social goal sharing technology to promote physical activity in children. MoHE 2016;5(2):-. [doi: 10.15282/mohe.v5i2.115]
19. Lau PW, Lau EY, Wong DP, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011 Jul 13;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
20. Druin A. The role of children in the design of new technology. Behav Inf Technol 2002;21(1):1-25. [doi: 10.1080/01449290110108659]
21. National Institute for Health and Care Excellence. National Institute for Health and Care Excellence. 2007. Behaviour change: general approaches URL: https://www.nice.org.uk/Guidance/ph6 [WebCite Cache ID 73BaRocnq]

22. Welk GJ. The youth physical activity promotion model: a conceptual bridge between theory and practice. Quest 1999;51(1):5-23. [doi: 10.1080/00336297.1999.10484297]
23. Greene S, Hogan D. Exploring Children's Views through Focus Groups. In: Researching Children's Experience. London: SAGE Publications Ltd; 2005.
24. Rita SD, Kenneth JD. Educational Leadership. 1979. Learning Styles/Teaching Styles: Should They... Can They... Be Matched? URL: http://www.ascd.org/ASCD/pdf/journals/ed_lead/el_197901_dunn.pdf [accessed 2018-11-29] [WebCite Cache ID 74Hkllyv3]
25. Lupton D. The quantified self. Cambridge: United Kingdom : Polity; 2016:68-183.
26. Hassenzahl M, Laschke M, Praest J. On the stories activity trackers tell. 2016 Presented at: UbiComp '16 Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; September 12 - 16, 2016; Heidelberg, Germany p. 582-587. [doi: 10.1145/2968219.2968325]
27. Li I, Anind KD, Forlizzi J. Understanding my data, myself: supporting self-reflection with ubicomp technologies. 2011 Presented at: UbiComp '11 Proceedings of the 13th international conference on Ubiquitous computing; September 17 - 21, 2011; Beijing, China p. 405-414. [doi: 10.1145/2030112.2030166]
28. Munson SA, Consolvo S. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. 2012 Presented at: 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops; 21-24 May 2012; San Diego, CA, USA. [doi: 10.4108/icst.pervasivehealth.2012.248691]
29. Kirk D, Sellen A. On human remains: values and practice in the home archiving of cherished objects. ACM Trans Comput-Hum Interact 2010;17(3):1-43. [doi: 10.1145/1806923.1806924]
30. Golsteijn C, Hoven EV, Frohlich D, Sellen A. Towards a More Cherishable Digital Object. 2012 Presented at: Proceedings of the Designing Interactive Systems Conference; 11-15 June, 2012; Newcastle upon Tyne, United Kingdom p. 655-664. [doi: 10.1145/2317956.2318054]
31. Ananthanarayan S, Siek K, Eisenberg M. A Craft Approach to Health Awareness in Children. 2016 Presented at: Proceedings of the 2016 ACM Conference on Designing Interactive Systems; June 04 - 08, 2016; Brisbane, QLD, Australia p. 724-735. [doi: 10.1145/2901790.2901888]
32. Gillet A, Sanner M, Stoffler D, Olson A. Tangible interfaces for structural molecular biology. Structure 2005 Mar;13(3):483-491 [FREE Full text] [doi: 10.1016/j.str.2005.01.009] [Medline: 15766549]
33. Mueller S, Mohr T, Guenther K, Frohnhofen J, Baudisch P. FaBrickation: fast 3D printing of functional objects by integrating construction kit building blocks. 2014 May Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014. ACM; 2014; Toronto, Canada. [doi: 10.1145/2556288.2557005]
34. Khot RA, Mueller FF, Hjorth L. SweatAtoms: materializing physical activity. 2013 Presented at: Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death; September 30 - October 01, 2013; Melbourne, Australia. [doi: 10.1145/2513002.2513012]
35. Khot RA, Pennings R, Mueller FF. EdiPulse: Turning Physical Activity Into Chocolates. 2015 Presented at: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems; April 18 - 23, 2015; Seoul, Republic of Korea. [doi: 10.1145/2702613.2725436]
36. Stusak S, Tabard A, Sauka F, Khot RA, Butz A. Activity sculptures: exploring the impact of physical visualizations on running activity. IEEE Trans Vis Comput Graph 2014;20(12):2201-2210. [doi: 10.1109/TVCG.2014.2352953]
37. Jansen Y, Dragicevic P, Fekete JD. Evaluating the efficiency of physical visualizations. 2013 May 02 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 02, 2013; Paris, France. [doi: 10.1145/2470654.2481359]
38. Khot RA. Exploring material representations of physical activity. 2014 Presented at: Proceedings of the 2014 companion publication on Designing interactive systems; June 21 - 25, 2014; Vancouver, BC, Canada p. 177-180. [doi: 10.1145/2598784.2598792]
39. Price S, Rogers Y, Scaife M, Stanton D, Neale H. Using ‘tangibles’ to promote novel forms of playful learning. Interact Comput 2003;15(2):169-185. [doi: 10.1016/S0953-5438(03)00006-7]
40. Rogers Y, Scaife M, Gabrielli S, Smith H, Harris E. A conceptual framework for mixed reality environments: designing novel learning activities for young children. Presence: Teleoperators Virtual Environ 2002;11(6):677-686. [doi: 10.1162/105474602321050776]
41. Locke E, Shaw KN, Saarj LM, Latham GP. Goal setting and task performance: 1969-1980. Psychological bulletin 1980;90(1):125-152. [doi: 10.1037/0033-2909.90.1.125]
42. Christian D, Todd C, Hill R, Rance J, Mackintosh K, Stratton G, et al. Active children through incentive vouchers ? evaluation (ACTIVE): a mixed-method feasibility study. BMC Public Health 2016;16(1):890. [doi: 10.1186/s12889-016-3381-6]
43. Finkelstein E, Tan YT, Malhotra R, Lee CF, Goh SS, Saw SM. A cluster randomized Controlled trial of an incentive-based outdoor physical activity program. J Pediatr 2013 Jul;163(1):167-172. [doi: 10.1016/j.jpeds.2013.01.009] [Medline: 23415616]
44. Hobbs G, Vignoles A. Is children’s free school meal ‘eligibility’ a good proxy for family income? Br Educ Res J 2010;36(4):673-690. [doi: 10.1080/01411920903083111]

45. Gibson F. Conducting focus groups with children and young people: strategies for success. J Res Nurs 2007;12(5):473-483. [doi: 10.1177/1744987107079791]
46. Liamputtong P. Researching the vulnerable: A guide to sensitive research methods. New York: Sage Publications; 2006.
47. Mackintosh K, Knowles ZR, Ridgers ND, Fairclough SJ. Using formative research to develop CHANGE!: a curriculum-based physical activity promoting intervention. BMC Public Health 2011;11:831. [doi: 10.1186/1471-2458-11-831]
48. Morgan M, Gibbs S, Maxwell K, Britten N. Hearing children's voices: methodological issues in conducting focus groups with children aged 7-11 years. Qual Res 2002;2(1):5-20. [doi: 10.1177/1468794102002001636]
49. Hill M, Laybourn A, Borland M. Engaging with primary-aged children about their emotions and well-being: methodological considerations. Child Soc 1996;10(2):129-144. [doi: 10.1111/j.1099-0860.1996.tb00463.x]
50. Kennedy C, Kools S, Krueger R. Methodological considerations in children's focus groups. Nurs Res 2001;50(3):184-187. [Medline: 11393641]
51. Dilorio C, Hockenberry-Eaton M, Maibach E, Rivero T. Focus groups: an interview method for nursing research. J Neurosci Nurs 1994;26(3):175-180. [Medline: 7963824]
52. Gibson J. Interviews and Focus Groups With Children: Methods That Match Children's Developing Competencies. J Fam Theory Rev 2012 May 31;4(2):148-159. [doi: 10.1111/j.1756-2589.2012.00119.x]
53. Clarke V, Braun V. Teaching thematic analysis: overcoming challenges and developing strategies for effective learning. The psychologist 2013;26(2):120-123.
54. Knowles ZR, Parnell D, Stratton G, Ridgers ND. Learning from the experts: exploring playground experience and activities using a write and draw technique. J Phys Act Health 2013 Mar;10(3):406-415. [Medline: 22820462]
55. Krane V, Andersen MB, Strean WB. Issues of Qualitative Research Methods and Presentation. J Sport Exerc Psychol 1997;19(2):213-218. [doi: 10.1123/jsep.19.2.213]
56. Ridgers ND, Knowles ZR, Sayers J. Encouraging play in the natural environment: a child-focused case study of Forest School. Children's Geographies 2012;10(1):49-65. [doi: 10.1080/14733285.2011.638176]
57. Hardy L, Jones JG, Gould D. Understanding Psychological Preparation for Sport: Theory and Practice of Elite Performers. Chichester: Wiley; 1996:364.
58. Sport Wales. Sport Wales. 5x60 URL: http://sport.wales/community-sport/education/5x60.aspx [accessed 2017-04-25] [WebCite Cache ID 73Bbuf9Fy]
59. Boddy LM, Knowles ZR, Davies IG, Warburton GL, Mackintosh KA, Houghton L, et al. Using formative research to develop the healthy eating component of the CHANGE! school-based curriculum intervention. BMC Public Health 2012 Aug 29;12:710 [FREE Full text] [doi: 10.1186/1471-2458-12-710] [Medline: 22931457]
60. Uttal D, O'Doherty K. Comprehending and learning from ‘visualizations’: a developmental perspective. In: Visualization: Theory and Practice in Science Education. Netherlands: Springer; 2008:53-72.
61. Piaget J, Margaret C. The origins of intelligence in children. In: International Universities Press New York. New York, NY: W W Norton & Co; 1952:424.
62. Marshall P. Do tangible interfaces enhance learning? 2007 Presented at: Proceedings of the 1st international conference on Tangible and embedded interaction; February 15 - 17, 2007; Baton Rouge, Louisiana p. 163-170. [doi: 10.1145/1226969.1227004]
63. Bara F, Gentaz E, Cole P, Sprenger-Charolles L. The visuo-haptic and haptic exploration of letters increases the kindergarten-children's of understanding of the alphabetic principle. Cogn Dev 2004;19(3):433-449. [doi: 10.1016/j.cogdev.2004.05.003]
64. Husmann PR, O'Loughlin VD. Another nail in the coffin for learning styles? Disparities among undergraduate anatomy students' study strategies, class performance, and reported VARK learning styles. Anat Sci Educ 2018 Mar 13:- Epub ahead of print(forthcoming). [doi: 10.1002/ase.1777] [Medline: 29533532]
65. Papanagnou D, Serrano A, Barkley K, Chandra S, Governatori N, Piela N, et al. Does tailoring instructional style to a medical student's self-perceived learning style improve performance when teaching intravenous catheter placement? A randomized controlled study. BMC Med Educ 2016 Aug 12;16(1):205 [FREE Full text] [doi: 10.1186/s12909-016-0720-3] [Medline: 27520578]
66. Riener C, Willingham D. The myth of learning styles. Change: The magazine of higher learning 2010;42(5):32-35.
67. Salvy S, Roemmich JN, Bowker JC, Romero ND, Stadler PJ, Epstein LH. Effect of peers and friends on youth physical activity and motivation to be physically active. J Pediatr Psychol 2009 Mar;34(2):217-225 [FREE Full text] [doi: 10.1093/jpepsy/jsn071] [Medline: 18617572]
68. Jago R, Page AS, Cooper AR. Friends and physical activity during the transition from primary to secondary school. Med Sci Sports Exerc 2012 Jan;44(1):111-117. [doi: 10.1249/MSS.0b013e318229df6e] [Medline: 21697746]
69. Salvy S, de la Haye K, Bowker JC, Hermans RC. Influence of peers and friends on children's and adolescents' eating and activity behaviors. Physiol Behav 2012 Jun 06;106(3):369-378 [FREE Full text] [doi: 10.1016/j.physbeh.2012.03.022] [Medline: 22480733]
70. Barkley JE, Salvy SJ, Sanders GJ, Dey S, Von Carlowitz KP, Williamson ML. Peer influence and physical activity behavior in young children: an experimental study. J Phys Act Health 2014;11(2):404-409.

71. Pearce M, Page AS, Griffin TP, Cooper AR. Who children spend time with after school: associations with objectively recorded indoor and outdoor physical activity. Int J Behav Nutr Phys Act 2014 Mar 30;11(1):45 [FREE Full text] [doi: 10.1186/1479-5868-11-45] [Medline: 24679149]
72. INEOS AG. 2013. Is competition good for kids? URL: https://www.ineos.com/inch-magazine/articles/issue-5/debate/ [accessed 2018-11-29] [WebCite Cache ID 74HoqQdN8]
73. Fry MD, Duda JL. A developmental examination of children's understanding of effort and ability in the physical and academic domains. Res Q Exerc Sport 1997 Dec;68(4):331-344. [doi: 10.1080/02701367.1997.10608014] [Medline: 9421845]
74. Jacobs JE, Klaczynski PA. The development of judgment and decision making during childhood and adolescence. Curr Dir Psychol Sci 2002;11(4):145-149. [doi: 10.1111/1467-8721.00188]
75. Benenson JF, Antonellis TJ, Cotton BJ, Noddin KE, Campbell KA. Sex differences in children's formation of exclusionary alliances under scarce resource conditions. Animal Behav 2008;76(2):497-505. [doi: 10.1016/j.anbehav.2008.01.027]
76. Björkqvist K. Sex differences in physical, verbal, and indirect aggression: a review of recent research. Sex Roles 1994;30(3):177-188. [doi: 10.1007/BF01420988]
77. Lin JJ, Mamykina L, Lindtner S, Delajoux G, Strub HB. Fish'n'Steps: encouraging physical activity with an interactive computer game. In: UbiComp 2006: Ubiquitous Computing. Berlin: Springer; 2006:261-278.
78. Hardman CA, Horne PJ, Fergus Lowe C. Effects of rewards, peer-modelling and pedometer targets on children's physical activity: a school-based intervention study. Psychol Health 2011;26(1):3-21. [doi: 10.1080/08870440903318119]
79. Strohacker K, Galarraga O, Williams DM. The impact of incentives on exercise behavior: a systematic review of randomized controlled trials. Ann Behav Med 2014;48(1):92-99. [doi: 10.1007/s12160-013-9577-4]
80. Khot RA. Exertion Games Lab. Melbourne, VIC 3000 Australia: RMIT University; 2016. Understanding Material Representations of Physical Activity URL: http://exertiongameslab.org/wp-content/uploads/2016/12/rohit_ashok_khot_phd.pdf [accessed 2018-11-29] [WebCite Cache ID 74HpBH9cx]
81. 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]
82. Davis WD. The interactive effects of goal orientation and feedback specificity on task performance. Hum Perform 2005;18(4):409-426. [doi: 10.1207/s15327043hup1804_7]
83. Stusak S, Schwarz J, Butz A. Evaluating the Memorability of Physical Visualizations. 2015 Presented at: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18 - 23, 2015; Seoul, Republic of Korea p. 3247-3250. [doi: 10.1145/2702123.2702248]
84. Consolvo S, Klasnja P, McDonald DW, Avrahami D, Froehlich J, LeGrand L, et al. Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays. 2008 Presented at: Proceedings of the 10th international conference on Ubiquitous computing; September 21 - 24, 2008; Seoul, Korea p. 54-63. [doi: 10.1145/1409635.1409644]
85. Anderson I, Maitland J, Sherwood S, Barkhuus L, Chalmers M, Brown M, et al. Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. Mobile Netw Appl 2007;12(2):185-199. [doi: 10.1007/s11036-007-0011-7]
86. Trost SG, Morgan AM, Saunders R, Felton G, Ward DS, Pate RR. Children's understanding of the concept of physical activity. Pediatr Exerc Sci 2000;12(3):293-299. [doi: 10.1123/pes.12.3.293]
87. Dourish P. Where the action is: the foundations of embodied interaction. Cambridge, MA: MIT press; 2004:233.
88. Jacob RJ, Ishii H, Pangaro G, Patten J. A tangible interface for organizing information using a grid. 2002 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 20 - 25, 2002; Minneapolis, Minnesota, USA p. 339-346. [doi: 10.1145/503376.503437]
89. Klahr D, Triona LM, Williams C. Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. J Res Sci Teach 2007;44(1):183-203. [doi: 10.1002/tea.20152]

Abbreviations

LED: light-emitting diode
PAL: physical activity levels
MVPA: moderate-to-vigorous physical activity
SES: socioeconomic status
YPAPM: Youth Physical Activity Promotion Model

Edited by G Eysenbach; submitted 29.08.18; peer-reviewed by R Wilbur, L Becker; comments to author 03.10.18; revised version received 16.10.18; accepted 17.10.18; published 30.01.19.

Please cite as:
Crossley SGM, McNarry MA, Hudson J, Eslambolchilar P, Knowles Z, Mackintosh KA
Perceptions of Visualizing Physical Activity as a 3D-Printed Object: Formative Study
J Med Internet Res 2019;21(1):e12064
URL: http://www.jmir.org/2019/1/e12064/
doi:10.2196/12064
PMID:30698532

Review

Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration

Bhone Myint Kyaw[1], MBBS, PhD; Nakul Saxena[2], PhD; Pawel Posadzki[3], PhD; Jitka Vseteckova[4], PhD; Charoula Konstantia Nikolaou[5], PhD; Pradeep Paul George[2], BSMS, MSc; Ushashree Divakar[3], MSc; Italo Masiello[6,7], PhD; Andrzej A Kononowicz[8], PhD; Nabil Zary[9,10,11], MD, PhD; Lorainne Tudor Car[1,12], MD, MSc, PhD

[1]Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[2]Health Services and Outcomes Research, National Healthcare Group Singapore, Singapore, Singapore
[3]Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[4]Faculty of Wellbeing, Education and Language Studies, The Open University, Milton Keynes, United Kingdom
[5]Centre de Philosophie du Proit, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium
[6]Department of Clinical Science and Education, Karolinska Institutet, Stockholm, Sweden
[7]Faculty of Social Sciences, Linnaeus University, Växjö, Sweden
[8]Department of Bioinformatics and Telemedicine, Jagiellonian University Medical College, Krakow, Poland
[9]Games for Health Innovations Centre, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
[10]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden
[11]International Medical Simulation Centre, Mohammed VI University of Health Sciences, Casablanca, Morocco
[12]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

Corresponding Author:
Lorainne Tudor Car, MD, MSc, PhD
Family Medicine and Primary Care
Lee Kong Chian School of Medicine
Nanyang Technological University Singapore
11 Mandalay Road
Singapore,
Singapore
Phone: 65 69041258
Email: lorainne.tudor.car@ntu.edu.sg

Abstract

Background: Virtual reality (VR) is a technology that allows the user to explore and manipulate computer-generated real or artificial three-dimensional multimedia sensory environments in real time to gain practical knowledge that can be used in clinical practice.

Objective: The aim of this systematic review was to evaluate the effectiveness of VR for educating health professionals and improving their knowledge, cognitive skills, attitudes, and satisfaction.

Methods: We performed a systematic review of the effectiveness of VR in pre- and postregistration health professions education following the gold standard Cochrane methodology. We searched 7 databases from the year 1990 to August 2017. No language restrictions were applied. We included randomized controlled trials and cluster-randomized trials. We independently selected studies, extracted data, and assessed risk of bias, and then, we compared the information in pairs. We contacted authors of the studies for additional information if necessary. All pooled analyses were based on random-effects models. We used the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach to rate the quality of the body of evidence.

Results: A total of 31 studies (2407 participants) were included. Meta-analysis of 8 studies found that VR slightly improves postintervention knowledge scores when compared with traditional learning (standardized mean difference [SMD]=0.44; 95% CI 0.18-0.69; I^2=49%; 603 participants; moderate certainty evidence) or other types of digital education such as online or offline digital education (SMD=0.43; 95% CI 0.07-0.79; I^2=78%; 608 participants [8 studies]; low certainty evidence). Another

meta-analysis of 4 studies found that VR improves health professionals' cognitive skills when compared with traditional learning (SMD=1.12; 95% CI 0.81-1.43; I^2=0%; 235 participants; large effect size; moderate certainty evidence). Two studies compared the effect of VR with other forms of digital education on skills, favoring the VR group (SMD=0.5; 95% CI 0.32-0.69; I^2=0%; 467 participants; moderate effect size; low certainty evidence). The findings for attitudes and satisfaction were mixed and inconclusive. None of the studies reported any patient-related outcomes, behavior change, as well as unintended or adverse effects of VR. Overall, the certainty of evidence according to the GRADE criteria ranged from low to moderate. We downgraded our certainty of evidence primarily because of the risk of bias and/or inconsistency.

Conclusions: We found evidence suggesting that VR improves postintervention knowledge and skills outcomes of health professionals when compared with traditional education or other types of digital education such as online or offline digital education. The findings on other outcomes are limited. Future research should evaluate the effectiveness of immersive and interactive forms of VR and evaluate other outcomes such as attitude, satisfaction, cost-effectiveness, and clinical practice or behavior change.

(J Med Internet Res 2019;21(1):e12959) doi:10.2196/12959

KEYWORDS

virtual reality; health professions education; randomized controlled trials; systematic review; meta-analysis

Introduction

Adequately trained health professionals are essential to ensure access to health services and to achieve universal health coverage [1]. In 2013, the World Health Organization (WHO) estimated a shortage of approximately 17.4 million health professionals worldwide [1]. The shortage and disproportionate distribution of health workers worldwide can be aggravated by the inadequacy of training programs (in terms of content, organization, and delivery) and experience needed to provide uniform health care services to all [2]. It has, therefore, become essential to generate strategies focused on scalable, efficient, and high-quality health professions education [3]. Increasingly, digital technology, with its pervasive use and relentless advancement, is seen as a promising source of effective and efficient health professions education and training systems [4].

Digital education (also known as eLearning) is the act of teaching and learning by means of digital technologies. It is an overarching term for an evolving multitude of educational approaches, concepts, methods, and technologies [5]. Digital education can include, but is not limited to, online and offline computer-based digital education, massive open online courses, virtual reality (VR), virtual patients, mobile learning, serious gaming and gamification, and psychomotor skills trainers [5]. A strong evidence base is needed to support effective use of these different digital modalities for health professions education. To this end, as part of an evidence synthesis series for digital health education, we focused on one of the digital education modalities, VR [6].

VR is a technology that allows the user to explore and manipulate computer-generated real or artificial three-dimensional (3D) multimedia sensory environments in real time. It allows for a first-person active learning experience through different levels of immersion; that is, a perception of the digital world as real and the ability to interact with objects and/or perform a series of actions in this digital world [7-9]. VR can be displayed with a variety of tools, including computer or mobile device screens, and VR rooms of head-mounted displays. VR rooms are projector-based immersive 3D visualization systems simulating real or virtual environments in a closed space and involve multiple users at the same time [10]. Head-mounted displays are placed over the user's head and provide an immersive 3D environmental experience for learning [7]. VR can also facilitate diverse forms of health professions education. For example, it is often used for designing 3D anatomical structure models, which can be toggled and zoomed into [11]. VR also enables the creation of virtual worlds or 3D environments with virtual representations of users, called avatars. Avatars in VR for health professions education can represent patients or health professionals. By enabling simulation, VR is highly conducive to clinical and surgical procedures-focused training.

We found several reviews focusing primarily on the development of technical skills as part of surgical and clinical procedures-focused training, mostly calling for more research on the topic [12-15]. However, VR also offers a range of other educational opportunities, such as development of cognitive, nontechnical competencies [13-18]. Our review addresses this gap in the existing literature by investigating the effectiveness of VR for health professions education.

Methods

Systematic Review

We adhered to the published protocol [6] and followed the Cochrane guidelines [19]. The review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [20]. For a detailed description of the methodology, please refer to the study by Car et al [5].

Study Selection

We included randomized and cluster-randomized controlled trials that compared any VR intervention with any control intervention, for the education of pre- or postregistration health professionals. We included health professionals with qualifications found in the Health Field of Education and Training (091) of the International Standard Classification of Education. VR interventions could be delivered as the only mode of education intervention or blended with traditional

learning (ie, blended learning). We included studies on VR for cognitive and nontechnical health professions education, including all VR delivery devices and levels of immersion. We included studies that reported VR as an intervention for healthcare professionals without the participant using any additional physical objects or devices such as probes or handles for psychomotor or technical skill development. We included studies that compared VR or blended learning with traditional learning, other types of digital educations, or another form of VR intervention.

We differentiated the following types of VR: 3D models, virtual patient or virtual health professional (VP or VHP) within VR and surgical simulation. Although we included studies including virtual patients in a VR, studies of virtual patient scenarios outside VR were excluded and are part of a separate review looking at virtual patients alone (simulation) [10]. We excluded studies of students and/or practitioners of traditional, alternative, and complementary medicine. We also excluded studies with cross-over design because of the likelihood of a carry-over effect.

We extracted data on the following primary outcomes:

- Learners' knowledge postintervention: Knowledge is defined as learners' factual or conceptual understanding measured using change between pre- and posttest scores.
- Learners' skills postintervention: Skills are defined as learners' ability to demonstrate a procedure or technique in an educational setting.
- Learners' attitudes postintervention toward new competences, clinical practice, or patients (eg, recognition of moral and ethical responsibilities toward patients): Attitude is defined as the tendency to respond positively or negatively toward the intervention.
- Learners' satisfaction postintervention with the learning intervention (eg, retention rates, dropout rates, and survey satisfaction scores): This can be defined as the level of approval when comparing the perceived performance of digital education with one's expectations.
- Change in learner's clinical practice or behavior (eg, reduced antibiotic prescriptions and improved clinical diagnosis): This can be defined as any changes in clinical practice after the intervention which results in improvement of the quality of care as well as the clinical outcomes.

We also extracted data on the following secondary outcomes:

- Cost and cost-effectiveness of the intervention
- Patient-related outcomes (eg, patient mortality, patient morbidity, and medication errors)

Data Sources, Collection, Analysis, and Risk of Bias Assessment

We developed a comprehensive search strategy for MEDLINE (Ovid), Embase (Elsevier), Cochrane Central Register of Controlled Trials (CENTRAL; Wiley), PsycINFO (Ovid), ERIC (Ovid), CINAHL (Ebsco), Web of Science Core Collection, and clinical trial registries (ClinicalTrial.gov and WHO ICTRP). Databases were searched from January 1990 to August 2017. The reason for selecting 1990 as the starting year for our search is that before this year, the use of computers and digital technologies was limited to very basic tasks. There were no language or publication restrictions (see Multimedia Appendix 1).

The search results from different bibliographic databases were combined in a single Endnote library, and duplicate records were removed. Four authors (BMK, NS, JV, and CKN) independently screened the search results and assessed full-text studies for inclusion. Any disagreements were resolved through discussion between the authors. Study authors were contacted for unclear or missing information.

Five reviewers (BMK, NS, JV, CKN, and UD) independently extracted data using a structured data extraction form. Disagreements between review authors were resolved by discussion. We extracted data on the participants, interventions, comparators, and outcomes. If studies had multiple arms, we compared the most interactive intervention arm with the least interactive control arm.

Two reviewers (BMK and NS) independently assessed the risk of bias for randomized controlled trials using the Cochrane *risk of bias* tool [19,21], which included the following domains: random sequence generation, allocation concealment, blinding of outcome assessors, completeness of outcome data, and selective outcome reporting. We also assessed the following additional sources of bias: baseline imbalance and inappropriate administration of an intervention as recommended by the Cochrane Handbook for Systematic Reviews of Interventions [21]. Studies were judged at high risk of bias if there was a high risk of bias for 1 or more key domains and at unclear risk of bias if they had an unclear risk of bias for at least 2 domains.

Data Synthesis and Analysis

Studies were grouped by outcome and comparison. Comparators included traditional education, other forms of digital education, and other types of VR. We included postintervention outcome data in our review for the sake of consistency as this was the most commonly reported form of findings in the included studies. For continuous outcomes, we summarized the standardized mean differences (SMDs) and associated 95% CIs across studies. We were unable to identify a clinically meaningful interpretation of SMDs specifically for digital education interventions. Therefore, in line with other evidence syntheses of educational research, we interpreted SMDs using the Cohen rule of thumb: <0.2 no effect, 0.2 to 0.5 small effect size, 0.5 to 0.8 medium effect size, and >0.80 a large effect size [22,22]. For dichotomous outcomes, we summarized relative risks and associated 95% CIs across studies. We employed the random-effects model in our meta-analysis. The I^2 statistic was employed to evaluate heterogeneity, with I^2<25%, 25% to 75%, and >75% to represent a low, moderate, and high degree of inconsistency, respectively. The meta-analysis was performed using Review Manager 5.3 (Cochrane Library Software, Oxford, UK). Where sufficient data were available, summary SMD and associated 95% CIs were estimated using random-effects meta-analysis [21].

We prepared *Summary of Findings* tables to present a summary of the results and a judgment on the quality of the evidence by using Grading of Recommendations, Assessment, Development

and Evaluations methodology. We presented the findings that we were unable to pool, because of lack of data or high heterogeneity, in the form of narrative synthesis.

Results

Results of the Search

The searches identified 30,532 unique references; of these, 31 studies (33 reports; 2407 participants) fulfilled the inclusion criteria [11,24-53] (see Figure 1).

Characteristics of Included Studies

All included studies were conducted in high-income countries. Moreover, 21 studies included only preregistration health professionals. A range of VR educational interventions were evaluated, including 3D models, VP or VHP within virtual worlds, and VR surgical stimulations. Control group interventions ranged from traditional learning (eg, lectures and textbooks) to other digital education interventions (online and offline) and other forms of VR (eg, with limited functions, noninteractivity, or nontutored support; see Multimedia Appendix 2). Although they met the inclusion criteria, some studies did not provide comparable outcome data. Out of the 24 studies assessing knowledge, 1 did not provide any comparable data to estimate the effect of the intervention [44]. Likewise, 2 out of 12 studies assessing skills [29,49], 4 out of 8 studies assessing attitude [38,40,46,50], and 8 out of 12 studies assessing satisfaction [24,29,33,37,48,49,52,53] did not provide comparable data.

Risk of Bias

Overall, studies were judged at unclear or high risk of bias (see Figure 2). Most studies lacked information on randomization, allocation concealment, and participants' baseline characteristics. Studies were mostly at low risk of bias for blinding of outcome assessment as they provided detailed information on blinding of outcome measures and/or used predetermined assessment tools (multiple choice questions, survey, etc). We judged the studies to be at low risk of detection bias in comparison with traditional education as blinding of participants was impossible because of the use of automated or formalized outcome measurement instruments. However, most of these instruments lacked information on validation. Most studies were judged to be at low risk of attrition and selection bias. Overall, 6 studies were judged at high risk of bias because of reported significant baseline differences in participant characteristics or incomplete outcome data.

Figure 1. Study flow diagram. RCT: randomized controlled trial.

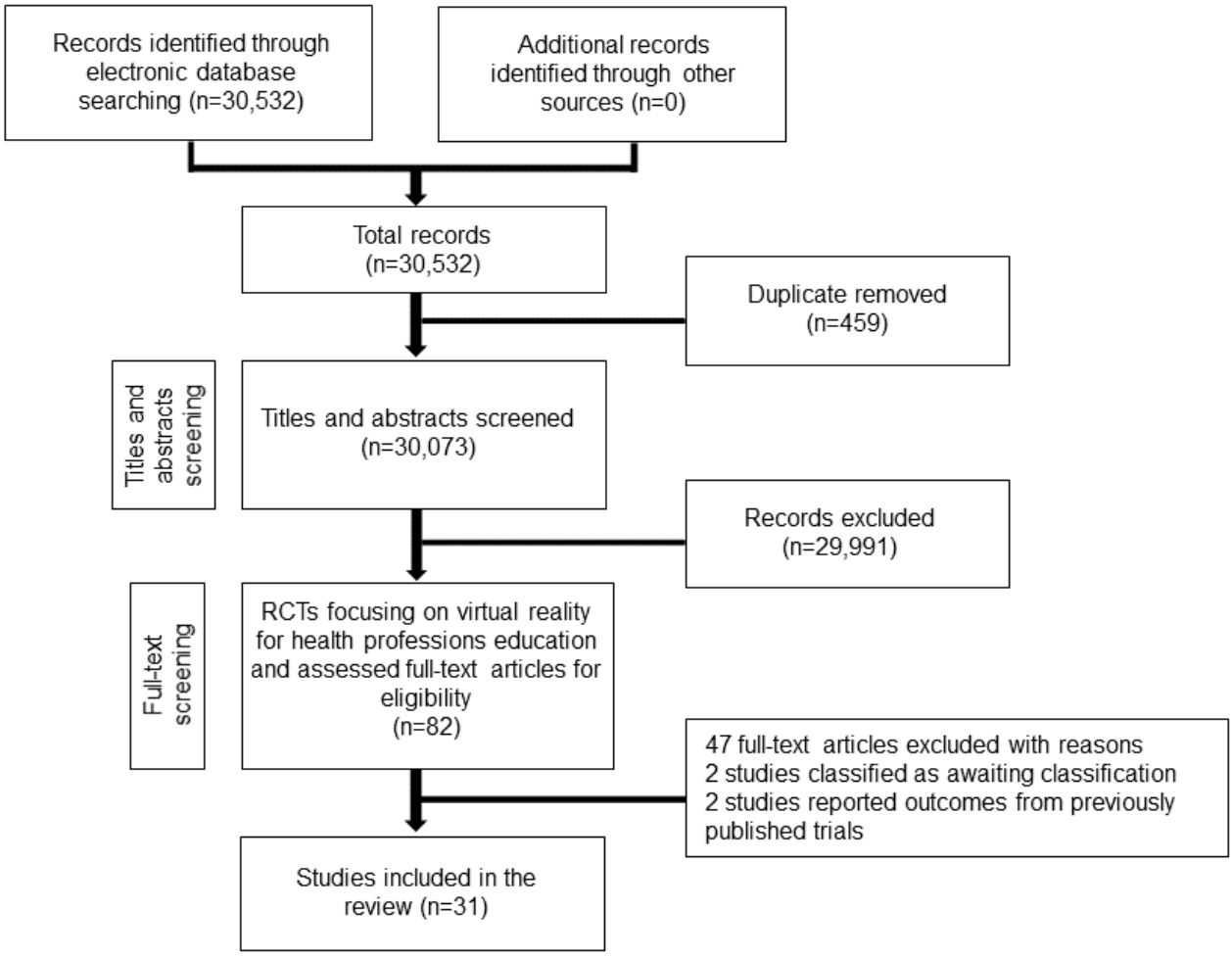
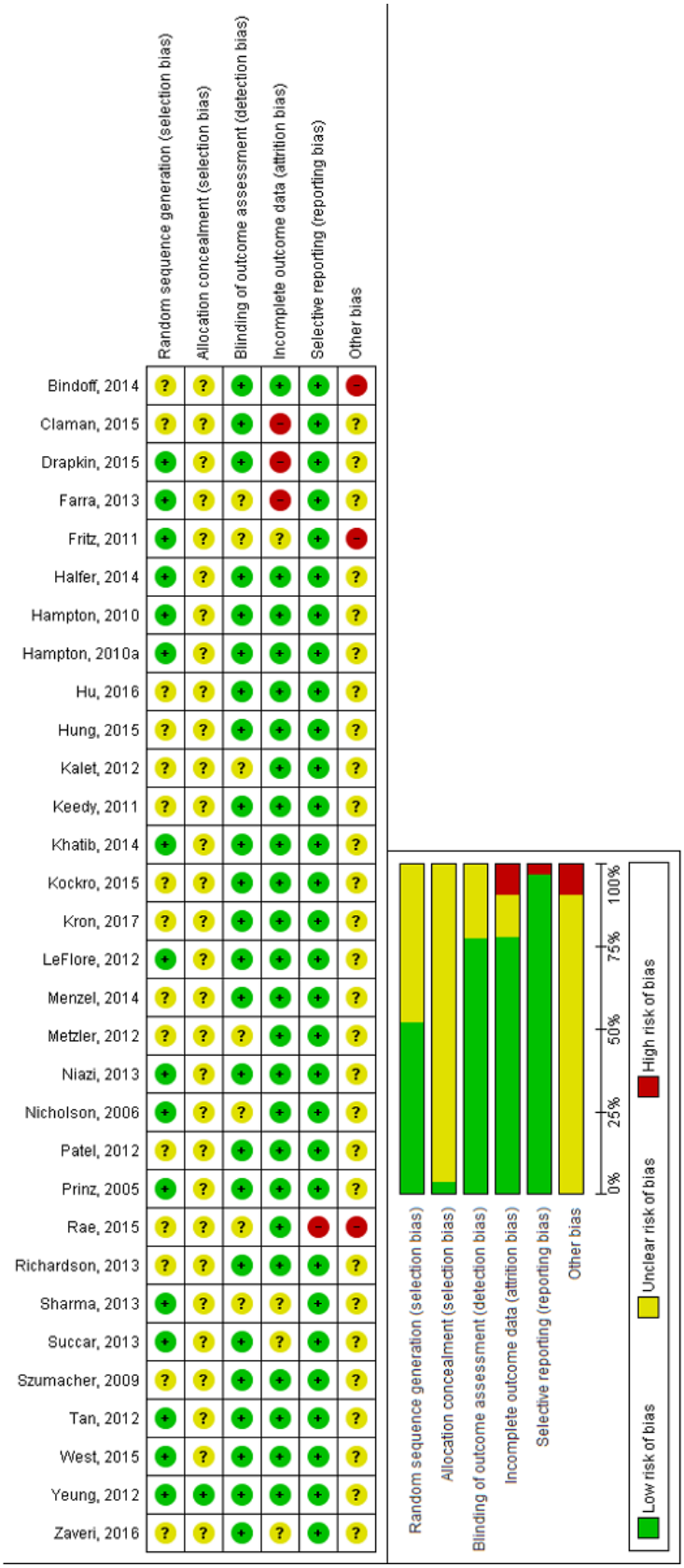


Figure 2. Risk of bias graph and summary.



Primary Outcomes

Knowledge Outcome

A total of 24 studies (1757 participants) [11,24,26-28,30-32,34-37,39,41-44,46-48,50-53] assessed knowledge as the primary outcome. Of them, 6 studies focused on postregistration health professionals [30,42,46,47,50,53] and all others focused on preregistration health professionals.

The effectiveness of VR interventions was compared with traditional learning (via two-dimensional [2D] images,

textbooks, and lectures) in 9 studies (659 participants) [24,30,32,36,39,43,47,48,52] (Table 1). Overall, studies suggested a slight improvement in knowledge with VR compared with traditional learning (SMD=0.44; 95% CI 0.18-0.69; I^2=49%; 603 participants [8 studies]; moderate certainty evidence; see Figure 3).

A total of 10 studies (812 participants) compared VR with other forms of digital education (comprising 2D images on a screen, simple videos, or Web-based teaching) [11,27,28,35,37,41,44,46,50,53] (see Table 2). The overall pooled estimate of 8 studies that compared different types of VR (such as computer 3D model and virtual world) with different controls (ie, computer-based 2D learning or online module or video-based learning) reported higher postintervention knowledge scores in the intervention groups over the control groups (SMD=0.43; 95% CI 0.07-0.79; I^2=78%; 608 participants; low certainty evidence; see Figure 3). Additionally, 4 studies compared 3D models with different levels of interactivity (243 participants) [26,34,42,51]. Models with higher interactivity were associated with greater improvements in knowledge than those with less interactivity. The overall pooled estimate of the 4 studies reported higher postintervention knowledge score in the intervention groups with higher interactivity compared with the less interactive controls (SMD=0.60; 95% CI 0.05-1.14; I^2=66%; moderate effect size; low certainty evidence; see Figure 3). A total of 3 studies could not be included in the meta-analysis: 1 study lacked data [44], whereas the other 2 studies reported a mean change score, favoring the VR group [36] or other digital education intervention [53].

Table 1. Summary of findings table: virtual reality compared with traditional learning.

Outcomes[a]	Illustrative comparative risks (95% CI)	Participants (n)	Studies (n)	Quality of evidence (GRADE[b])	Comments
Postintervention knowledge scores: measured via MCQs[c] or quiz. Follow-up: immediate postintervention only	The mean knowledge score in the intervention group was 0.44 SDs higher (0.18 to 0.69 higher) than the mean score in the traditional learning group	603	8	Moderate[d]	1 study [36] reported mean change scores within the group, and hence, the study data were excluded from the pooled analysis
Postintervention skill scores: measured via survey and OSCE[e]. Follow-up duration: immediate postintervention only	The mean skill score in the intervention group was 1.12 SDs higher (0.81 to 1.43 higher) than the mean score in the traditional learning group	235	4	Moderate[d]	3 studies were excluded from the analysis as 1 study reported incomplete outcome data [29], 1 study assessed mixed outcomes [36], and 1 study reported self-reported outcome data [24]
Postintervention attitude scores: measured via survey. Follow-up duration: immediate postintervention only	The mean attitudinal score in the intervention group was 0.19 SDs higher (−0.35 lower to 0.73 higher) than the mean score in the traditional learning group	83	2	Moderate[d]	N/A[f]
Postintervention satisfaction scores: measured via survey. Follow-up duration: immediate postintervention only	Not estimable	100	1	Low[d,g]	5 studies [24,29,33,48,52] reported incomplete outcome data or lacked comparable data. Therefore, these studies were excluded from the analysis.

[a]Patient or population: health professionals; settings: universities and hospitals; intervention: virtual reality; comparison: traditional learning (face-to-face lecture, textbooks, etc).

[b]GRADE (Grading of Recommendations, Assessment, Development and Evaluations) Working Group grades of evidence. High quality: further research is very unlikely to change our confidence in the estimate of effect; moderate quality: further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate; low quality: further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate; and very low quality: we are very uncertain about the estimate.

[c]MCQs: multiple choice questions.

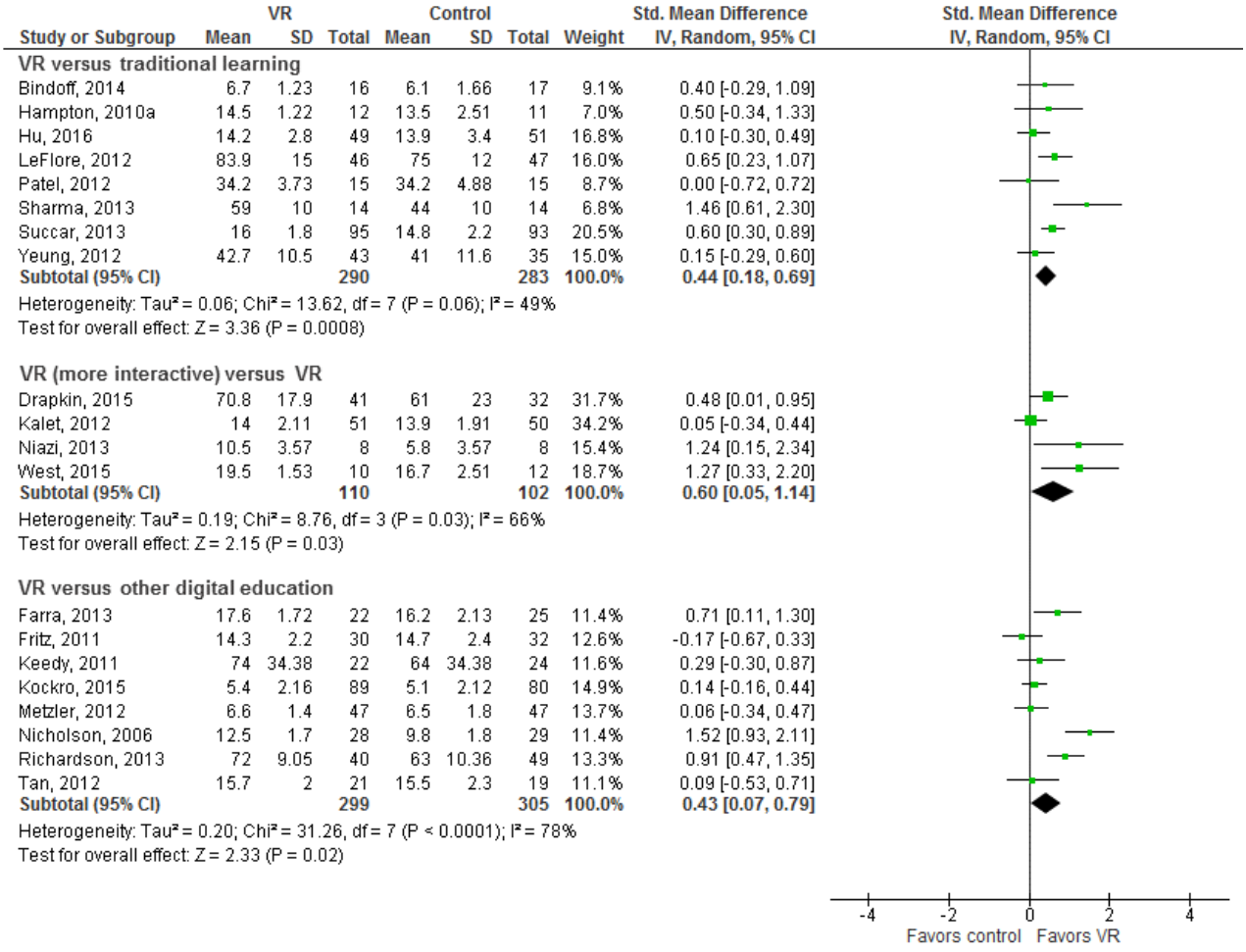
[d]Downgraded by 1 level for study limitations: the risk of bias was unclear or high in most included studies (−1).

[e]OSCE: objective structured clinical examination.

[f]N/A: not applicable.

[g]Downgraded as results were obtained from a single small study (−1).

Figure 3. Forest plot for the knowledge outcome (postintervention). df: degrees of freedom; IV: interval variable; random: random effects model; VR: virtual reality.



Skill Outcome

A total of 12 studies (1011 participants) assessed skills as an outcome [24,29,33,34,36,38,39,42,43,45,49,53]. Of which, 7 studies compared VR-based interventions with traditional learning (comprising paper- or textbook-based education and didactic lectures; 354 participants) [24,29,33,36,39,43,45], and the overall pooled estimate of 4 studies showed a large improvement in postintervention cognitive skill scores in the intervention groups compared with the controls (SMD=1.12; 95% CI 0.81-1.43; I^2=0%, 235 participants; large effect size; moderate certainty evidence; see Figure 4 and Table 1). Additionally, 3 studies compared the effectiveness of different types of VR on cognitive skills acquisition (190 participants) [34,42,49]. We were able to pool the findings from 2 studies favoring more interactive VR (SMD=0.57; 95% CI 0.19-0.94; I^2=0%; moderate effect size; low certainty evidence). Two studies compared VR with other forms of digital education on skills, favoring the VR group (SMD=0.5; 95% CI 0.32-0.69; I^2=0%; 467 participants; moderate effect size; low certainty evidence; see Table 2). A total of 4 studies could not be included in the meta-analysis: 2 studies reported incomplete outcome data [29,49], 1 study assessed mixed outcomes [36], and 1 study reported self-reported outcome data [24].

Attitude Outcome

A total of 8 studies (762 participants) [25,30,31,38,40,43,46,50] assessed attitude as an outcome. Of them, 2 studies comparing VR-based interventions with traditional learning (small group teaching and didactic lectures; 83 participants) [30,43] reported no difference between the groups on postintervention attitude scores (SMD=0.19; 95% CI−0.35 to 0.73; I^2=0%; moderate certainty evidence; see Table 1). One study compared blended learning with traditional learning (43 participants) [30] and reported higher postinterventional attitude score (large effect size) toward the intervention (SMD=1.11; 95% CI 0.46-1.75). Additionally, 5 studies (636 participants) [25,38,40,46,50] that compared VR with other digital education interventions reported that most of the studies had incomplete outcome data.

Table 2. Summary of findings table: virtual reality compared with other digital education interventions.

Outcomes[a]	Illustrative comparative risks (95% CI)	Participants (n)	Studies (n)	Quality of evidence (GRADE[b])	Comments
Postintervention knowledge score: measured via MCQs[c] and questionnaires. Follow-up duration: immediate postintervention to 6 months	The mean knowledge score in the intervention group was 0.43 SDs higher (0.07 to 0.79 higher) than the mean score in the other digital education interventions	608	8	Low[d,e]	1 study (32 participants) presented mean change score and favored VR group compared with the control group [53], and 1 study (172 participants) compared VR with computer-based video (2D) and presented incomplete outcome data [44]
Postintervention skills score: measured via scenario-based skills assessment. Follow-up duration: immediate postintervention only	The mean skill score in the intervention group was 0.5 SDs higher (0.32 to 0.69 higher) than the mean score in the other digital education interventions	467	2	Moderate[d]	N/A[f]
Postintervention attitude: measured via survey and questionnaire. Follow-up duration: immediate postintervention only.	Not estimable	21	1	Low[d,g]	4 studies [38,40,46,50] reported incomplete outcome data or lacked comparable data. Therefore, these studies were excluded from the analysis.
Postintervention satisfaction: measured via MCQs, survey, and questionnaire. Duration: immediate postintervention only	The mean satisfaction score in the intervention group was 0.2 SDs higher (−0.71 lower to 1.11 higher) than the mean score in the other digital education interventions	218	2	Low[d,e]	2 studies [37,53] reported incomplete outcome data or lacked comparable data. Therefore, these studies were excluded from the analysis.

[a]Patient or population: Health professionals; Settings: Universities and hospitals; Intervention: Virtual reality; Comparison: Other digital education interventions (such as online learning, computer-based video, etc).

[b]GRADE (Grading of Recommendations, Assessment, Development and Evaluations) Working Group grades of evidence. High quality: Further research is very unlikely to change our confidence in the estimate of effect; Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate; Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate; and Very low quality: We are very uncertain about the estimate.

[c]MCQs: multiple choice questions.

[d]Downgraded by 1 level for study limitations (−1): the risk of bias was unclear or high in most included studies.

[e]Downgraded by 1 level for inconsistency (−1): the heterogeneity between studies is high with large variations in effect and lack of overlap among confidence intervals.

[f]N/A: not applicable.

[g]Downgraded as results were obtained from a single small study (−1).

Satisfaction Outcome

A total of 12 studies (991 participants) [24,26,29,32,33,35,37,44,48,49,52,53] assessed satisfaction, mostly only for the intervention group. Only 4 studies compared satisfaction in the intervention and control groups and largely reported no difference between the study groups.

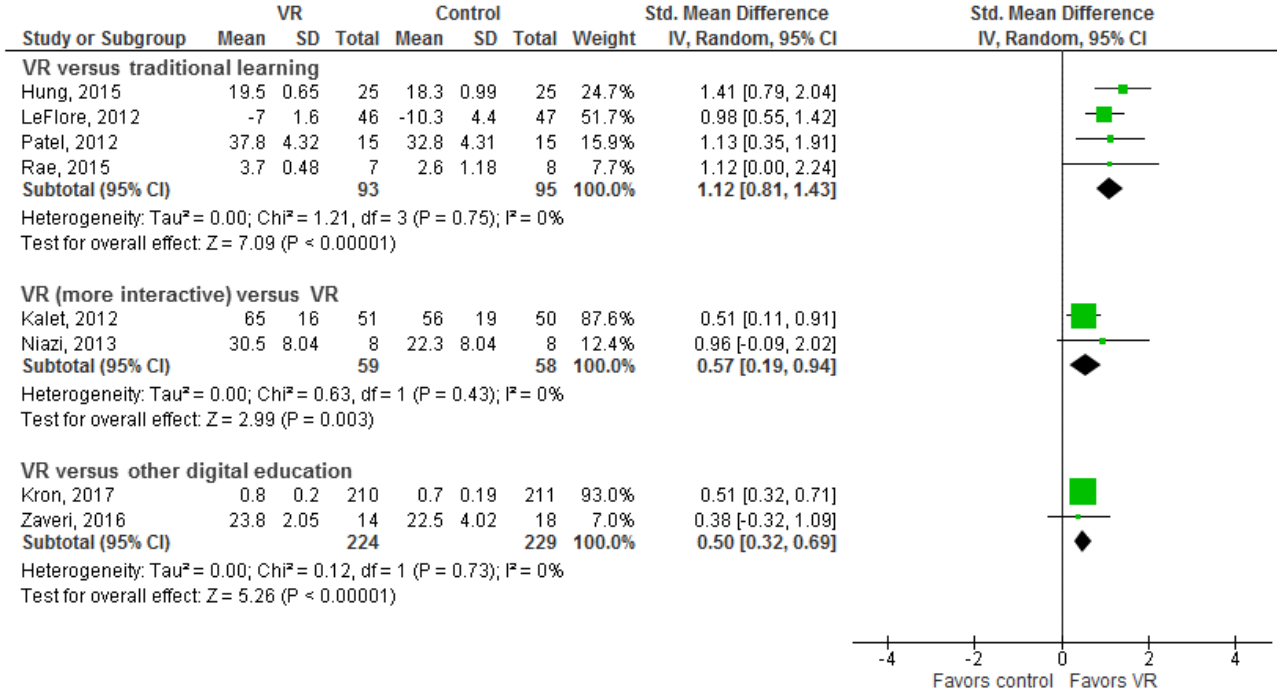
Secondary Outcomes

Halfer et al (30 participants) [29] assessed the use of VR versus traditional paper floor plans of the hospital to prepare nurses for wayfinding in a new hospital building. A cost analysis showed that a virtual hospital-based approach increased development costs but provided increased value during implementation by reducing staff time needed for practicing wayfinding skills. The paper describes that the real-world paper floor plan approach had a development cost of US $40,000 and the implementation cost was US $530,000, bringing the total cost to US $570,000. In comparison, the virtual world would cost US $220,000 for development and US $201,000 for implementation, bringing the total to US $421,000.

Zaveri et al (32 participants) [53] assessed the effectiveness of a VR module (Second Life, Linden Lab) in teaching preparation and management of sedation procedures, compared with online learning. Development of the module occurred over 2 years of interactions with a software consultant and utilized a US $40,000 grant to create VR scenarios. Cost of the control group (online training) was not presented, and hence, no formal comparison was made.

No information on patient-related outcomes, behavior change, and unintended or adverse effects of VR on the patient or the learner were reported in any of the studies.

Figure 4. Forest plot for the skills outcome (postintervention). df: degrees of freedom; IV: interval variable; random: random effects model; VR: virtual reality.



Discussion

Principal Findings

This systematic review assessed the effectiveness of VR interventions for health professions education. We found evidence showing a small improvement in knowledge and moderate-to-large improvement in skills in learners taking part in VR interventions compared with traditional or other forms of digital learning. Compared with less interactive interventions, more interactive VR interventions seem to moderately improve participants' knowledge and skills. The findings for attitude and satisfaction outcomes are inconclusive because of incomplete outcome data. None of the included studies reported any patient-related outcomes, behavior change, as well as unintended or adverse effects of the VR on the patients or the learners. Only 2 studies assessed the cost of setup and maintenance of the VR as a secondary outcome without making any formal comparisons.

Overall, the risk of bias for most studies was judged to be unclear (because of a lack of information), with some instances of potentially high risk of attrition, reporting, and other bias identified. The quality of the evidence ranges from low to moderate for knowledge, skills, attitude, and satisfaction outcomes because of the unclear and high risks of bias and inconsistency, that is, heterogeneity in study results as well as in types of participants, interventions, and outcome measurement instruments [54].

The fact that no included studies were published before 2005 suggests that VR is an emerging educational strategy, attracting increasing levels of interest. The included studies were mainly conducted among doctors, nurses, and students pursuing their medical degree. Limited studies on pharmacists, dentists, and other allied health professionals suggest more research is needed on the use of VR among these groups of health professionals. Additionally, the majority of interventions studied were not part of a regular curriculum and none of the studies mentioned the use of learning theories to design the VR-based intervention or develop clinical competencies. This is an important aspect of designing any curriculum, and hence, applicability of the included studies might only be limited to their current setting and may not be generalizable to other geographic or socioeconomic backgrounds. Furthermore, most studies evaluated participants' knowledge, and skills assessed may not translate directly into clinical competencies.

Although the included studies encompassed a range of participants and interventions, a lack of consistent methodological approach and studies conducted in any one health care discipline makes it difficult to draw any meaningful conclusions. There is also a distinct lack of data from low- or middle-income countries, which reduces applicability to those contexts that are most in need of innovative educational strategies. In addition, only 2 studies assessed the cost of setup and maintenance of the VR-based intervention, whereas none of the included studies assessed cost-effectiveness. Thus, no conclusions regarding costing and cost-effectiveness can be made at this point either. There was also a lack of information on patient-related outcomes, behavior change, and unintended or adverse effects of VR on the patients as well as the learner, which needs to be addressed.

Majority of the included studies assessed the effectiveness of nonimmersive VR, and there is a need to explore more on the effects of VR with different level of immersion as well as interactivity on the outcomes of interest. In our review, most of the studies assessing attitude and satisfaction outcomes reported incomplete outcome or incomparable outcome data,

and there is a need for primary studies focusing on these outcomes. Finally, there is a need to standardize the methods for reporting meaningful and the most accurate data on the outcomes as most of the included studies reported postintervention mean scores rather than change scores on the outcomes, which limits the accuracy of the findings for the reported outcomes.

Strengths and Limitations

Our review provides the most up-to-date evidence on the effectiveness of different types of VR in health professions education. We conducted a comprehensive search across different databases including gray literature sources and followed the Cochrane gold standard methodology while conducting this systematic review. Our review also has several limitations. The included studies largely reported postintervention data, so we could not calculate pre- to postintervention change or ascertain whether the intervention groups were matched at baseline for key characteristics and outcome measure scores. We were also unable to perform the prespecified subgroup analysis because of limited data from the primary studies.

Conclusions

As an emerging and versatile technology, VR has the potential to transform health professions education. Our findings show that when compared with traditional education or other types of digital education, such as online or offline digital education, VR may improve postintervention knowledge and skills. VR with higher interactivity showed more effectiveness compared with less interactive VR for postintervention knowledge and skill outcomes. Further research should evaluate the effectiveness of more immersive and interactive forms of VR in a variety of settings and evaluate outcomes such as attitude, satisfaction, untoward effects of VR, cost-effectiveness, and change in clinical practice or behavior.

Acknowledgments

This review was conducted in collaboration with the Health Workforce Department at WHO. The authors would also like to thank Mr Carl Gornitzki, Ms GunBrit Knutssön, and Mr Klas Moberg from the University Library, Karolinska Institutet, Sweden, for developing the search strategy and the peer reviewers for their comments. The authors would like to thank Associate Professor Josip Car for his valuable inputs throughout the drafting and revision of the review. The authors would like to thank Dr Parvati Dev for her valuable inputs from a content expert point of view. The authors would also like to thank Dr Gloria Law for her inputs on the classification of measurement instruments used to assess outcomes in the included studies. The authors would also like to thank Dr Ram Bajpai for his statistical advice. The authors gratefully acknowledge funding from the Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore: eLearning for health professionals education grant.

Authors' Contributions

LTC and NZ conceived the idea for the review. BMK, NS, and LTC wrote the review. LTC and PP provided methodological guidance, drafted some of the methodology-related sections, and critically revised the review. JV, PPG, IM, UD, AAK, CKN, and NZ provided comments on and edited the review.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE (Ovid) search strategy.

[PDF File (Adobe PDF File), 45KB - jmir_v21i1e12959_app1.pdf]

Multimedia Appendix 2

Characteristics of included studies.

[DOCX File, 44KB - jmir_v21i1e12959_app2.pdf]

References

1. World Health Organization. 2016 Mar 02. High-Level Commission on Health Employment and Economic Growth URL: https://www.who.int/hrh/com-heeg/en/ [accessed 2019-01-14] [WebCite Cache ID 75PcWOWNz]
2. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. Lancet 2010 Dec 04;376(9756):1923-1958. [doi: 10.1016/S0140-6736(10)61854-5] [Medline: 21112623]
3. Transformative scale up of health professional education: an effort to increase the numbers of health professionals and to strengthen their impact on population health. Geneva: World Health Organization; 2011. URL: http://apps.who.int/iris/handle/10665/70573 [accessed 2019-01-14] [WebCite Cache ID 75Pe1VnAx]

4. Crisp N, Gawanas B, Sharp I, Task Force for Scaling Up EducationTraining for Health Workers. Training the health workforce: scaling up, saving lives. Lancet 2008 Feb 23;371(9613):689-691. [doi: 10.1016/S0140-6736(08)60309-8] [Medline: 18295028]
5. Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, et al. Digital education for health professions: methods for overarching evidence syntheses. J Med Internet Res 2019:2018 (forthcoming). [doi: 10.2196/preprints.12913]
6. Saxena N, Kyaw B, Vseteckova J, Dev P, Paul P, Lim KT, et al. Virtual reality environments for health professional education. Cochrane Database Syst Rev 2016 Feb 09(2):CD012090 [FREE Full text] [doi: 10.1002/14651858.CD012090]
7. Mantovani F, Castelnuovo G, Gaggioli A, Riva G. Virtual reality training for health-care professionals. Cyberpsychol Behav 2003 Aug;6(4):389-395. [doi: 10.1089/109493103322278772] [Medline: 14511451]
8. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003-2009. Med Educ 2010 Jan;44(1):50-63. [doi: 10.1111/j.1365-2923.2009.03547.x] [Medline: 20078756]
9. Rasmussen K, Belisario JM, Wark PA, Molina JA, Loong SL, Cotic Z, et al. Offline eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014 Jun;4(1):010405 [FREE Full text] [doi: 10.7189/jogh.04.010405] [Medline: 24976964]
10. Andreatta PB, Maslowski E, Petty S, Shim W, Marsh M, Hall T, et al. Virtual reality triage training provides a viable solution for disaster-preparedness. Acad Emerg Med 2010 Aug;17(8):870-876 [FREE Full text] [doi: 10.1111/j.1553-2712.2010.00728.x] [Medline: 20670325]
11. Nicholson DT, Chalk C, Funnell WR, Daniel SJ. Can virtual reality improve anatomy education? A randomised controlled study of a computer-generated three-dimensional anatomical ear model. Med Educ 2006 Nov;40(11):1081-1087. [doi: 10.1111/j.1365-2929.2006.02611.x] [Medline: 17054617]
12. Gurusamy KS, Aggarwal R, Palanivelu L, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 2009 Jan 21(1):CD006575. [doi: 10.1002/14651858.CD006575.pub2] [Medline: 19160288]
13. Nagendran M, Gurusamy KS, Aggarwal R, Loizidou M, Davidson BR. Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 2013 Aug 27(8):CD006575. [doi: 10.1002/14651858.CD006575.pub3] [Medline: 23980026]
14. Piromchai P, Avery A, Laopaiboon M, Kennedy G, O'Leary S. Virtual reality training for improving the skills needed for performing surgery of the ear, nose or throat. Cochrane Database Syst Rev 2015 Sep 09(9):CD010198. [doi: 10.1002/14651858.CD010198.pub2] [Medline: 26352008]
15. Walsh CM, Sherlock ME, Ling SC, Carnahan H. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. Cochrane Database Syst Rev 2012 Jun 13(6):CD008237. [doi: 10.1002/14651858.CD008237.pub2] [Medline: 22696375]
16. Fritz PZ, Gray T, Flanagan B. Review of mannequin-based high-fidelity simulation in emergency medicine. Emerg Med Australas 2008 Feb;20(1):1-9. [doi: 10.1111/j.1742-6723.2007.01022.x] [Medline: 17999685]
17. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 2005 Jan;27(1):10-28. [doi: 10.1080/01421590500046924] [Medline: 16147767]
18. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. Acad Med 2003 Aug;78(8):783-788. [Medline: 12915366]
19. Cochrane Library, The Cochrane Collaboration. 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011] URL: http://crtha.iums.ac.ir/files/crtha/files/cochrane.pdf [accessed 2019-01-15] [WebCite Cache ID 75RKkvhhz]
20. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
21. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928. [doi: 10.1136/bmj.d5928] [Medline: 22008217]
22. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. J Am Med Assoc 2011 Sep 07;306(9):978-988. [doi: 10.1001/jama.2011.1234] [Medline: 21900138]
23. RevMan 5.3. Copenhagen: The Nordic Cochrane Centre: The Cochrane Collaboration; 2014 Jun 13. URL: https://community.cochrane.org/sites/default/files/uploads/inline-files/RevMan_5.3_User_Guide.pdf [accessed 2019-01-15] [WebCite Cache ID 75RLEAsgT]
24. Bindoff I, Ling T, Bereznicki L, Westbury J, Chalmers L, Peterson G, et al. A computer simulation of community pharmacy practice for educational use. Am J Pharm Educ 2014 Nov 15;78(9):168 [FREE Full text] [doi: 10.5688/ajpe789168] [Medline: 26056406]
25. Claman FL. The impact of multiuser virtual environments on student engagement. Nurse Educ Pract 2015 Jan;15(1):13-16. [doi: 10.1016/j.nepr.2014.11.006] [Medline: 25532889]
26. Drapkin ZA, Lindgren KA, Lopez MJ, Stabio ME. Development and assessment of a new 3D neuroanatomy teaching tool for MRI training. Anat Sci Educ 2015;8(6):502-509. [doi: 10.1002/ase.1509] [Medline: 25573020]

27. Farra S, Miller E, Timm N, Schafer J. Improved training for disasters using 3-D virtual reality simulation. West J Nurs Res 2013 May;35(5):655-671. [doi: 10.1177/0193945912471735] [Medline: 23287551]
28. Fritz D, Hu A, Wilson T, Ladak H, Haase P, Fung K. Long-term retention of a 3-dimensional educational computer model of the larynx: a follow-up study. Arch Otolaryngol Head Neck Surg 2011 Jun;137(6):598-603. [doi: 10.1001/archoto.2011.76] [Medline: 21690512]
29. Halfer D, Rosenheck M. Virtual education: is it effective for preparing nurses for a hospital move? J Nurs Adm 2014 Oct;44(10):535-540. [doi: 10.1097/NNA.0000000000000112] [Medline: 25280076]
30. Hampton BS, Sung VW. Improving medical student knowledge of female pelvic floor dysfunction and anatomy: a randomized trial. Am J Obstet Gynecol 2010 Jun;202(6):601.e1-601.e8. [doi: 10.1016/j.ajog.2009.08.038] [Medline: 20430356]
31. Hampton BS, Sung VW. A randomized trial to estimate the effect of an interactive computer trainer on resident knowledge of female pelvic floor dysfunction and anatomy. Female Pelvic Med Reconstr Surg 2010 Jul;16(4):224-228. [doi: 10.1097/SPV.0b013e3181ed3fb1] [Medline: 22453346]
32. Hu A, Shewokis PA, Ting K, Fung K. Motivation in computer-assisted instruction. Laryngoscope 2016 Aug;126 Suppl 6:S5-S13. [doi: 10.1002/lary.26040] [Medline: 27307270]
33. Hung P, Choi KS, Chiang VC. Using interactive computer simulation for teaching the proper use of personal protective equipment. Comput Inform Nurs 2015 Feb;33(2):49-57. [doi: 10.1097/CIN.0000000000000125] [Medline: 25521788]
34. Kalet AL, Song HS, Sarpel U, Schwartz R, Brenner J, Ark TK, et al. Just enough, but not too much interactivity leads to better clinical skills performance after a computer assisted learning module. Med Teach 2012 Aug;34(10):833-839 [FREE Full text] [doi: 10.3109/0142159X.2012.706727] [Medline: 22917265]
35. Keedy AW, Durack JC, Sandhu P, Chen EM, O'Sullivan PS, Breiman RS. Comparison of traditional methods with 3D computer models in the instruction of hepatobiliary anatomy. Anat Sci Educ 2011 Mar;4(2):84-91. [doi: 10.1002/ase.212] [Medline: 21412990]
36. Khatib M, Hald N, Brenton H, Barakat MF, Sarker SK, Standfield N, et al. Validation of open inguinal hernia repair simulation model: a randomized controlled educational trial. Am J Surg 2014 Aug;208(2):295-301. [doi: 10.1016/j.amjsurg.2013.12.007] [Medline: 24581993]
37. Kockro RA, Amaxopoulou C, Killeen T, Wagner W, Reisch R, Schwandt E, et al. Stereoscopic neuroanatomy lectures using a three-dimensional virtual reality environment. Ann Anat 2015 Sep;201:91-98. [doi: 10.1016/j.aanat.2015.05.006] [Medline: 26245861]
38. Kron FW, Fetters MD, Scerbo MW, White CB, Lypson ML, Padilla MA, et al. Using a computer simulation for teaching communication skills: a blinded multisite mixed methods randomized controlled trial. Patient Educ Couns 2017 Apr;100(4):748-759. [doi: 10.1016/j.pec.2016.10.024] [Medline: 27939846]
39. LeFlore JL, Anderson M, Zielke MA, Nelson KA, Thomas PE, Hardee G, et al. Can a virtual patient trainer teach student nurses how to save lives--teaching nursing students about pediatric respiratory diseases. Simul Healthc 2012 Feb;7(1):10-17. [doi: 10.1097/SIH.0b013e31823652de] [Medline: 22228285]
40. Menzel N, Willson LH, Doolen J. Effectiveness of a poverty simulation in Second Life®: changing nursing student attitudes toward poor people. Int J Nurs Educ Scholarsh 2014 Mar 11;11(1):39-45. [doi: 10.1515/ijnes-2013-0076] [Medline: 24615491]
41. Metzler R, Stein D, Tetzlaff R, Bruckner T, Meinzer HP, Büchler MW, et al. Teaching on three-dimensional presentation does not improve the understanding of according CT images: a randomized controlled study. Teach Learn Med 2012 Apr;24(2):140-148. [doi: 10.1080/10401334.2012.664963] [Medline: 22490095]
42. Niazi AU, Tait G, Carvalho JC, Chan VW. The use of an online three-dimensional model improves performance in ultrasound scanning of the spine: a randomized trial. Can J Anaesth 2013 May;60(5):458-464. [doi: 10.1007/s12630-013-9903-0] [Medline: 23400584]
43. Patel V, Aggarwal R, Osinibi E, Taylor D, Arora S, Darzi A. Operating room introduction for the novice. Am J Surg 2012 Feb;203(2):266-275. [doi: 10.1016/j.amjsurg.2011.03.003] [Medline: 21703594]
44. Prinz A, Bolz M, Findl O. Advantage of three dimensional animated teaching over traditional surgical videos for teaching ophthalmic surgery: a randomised study. Br J Ophthalmol 2005 Nov;89(11):1495-1499 [FREE Full text] [doi: 10.1136/bjo.2005.075077] [Medline: 16234460]
45. Rae AO, Khatib M, Sarker SK, Bello F. The effect of a computer based open surgery simulation of an inguinal hernia repair on the results of cognitive task analysis performance of surgical trainees: an educational trial. Br J Surg 2015;102(S1):40.
46. Richardson A, Bracegirdle L, McLachlan SI, Chapman SR. Use of a three-dimensional virtual environment to teach drug-receptor interactions. Am J Pharm Educ 2013 Feb 12;77(1):11 [FREE Full text] [doi: 10.5688/ajpe77111] [Medline: 23459131]
47. Sharma V, Chamos C, Valencia O, Meineri M, Fletcher SN. The impact of internet and simulation-based training on transoesophageal echocardiography learning in anaesthetic trainees: a prospective randomised study. Anaesthesia 2013 Jun;68(6):621-627 [FREE Full text] [doi: 10.1111/anae.12261] [Medline: 23594265]
48. Succar T, Zebington G, Billson F, Byth K, Barrie S, McCluskey P, et al. The impact of the Virtual Ophthalmology Clinic on medical students' learning: a randomised controlled trial. Eye (Lond) 2013 Oct;27(10):1151-1157 [FREE Full text] [doi: 10.1038/eye.2013.143] [Medline: 23867718]

49. Szumacher E, Harnett N, Warner S, Kelly V, Danjoux C, Barker R, et al. Effectiveness of educational intervention on the congruence of prostate and rectal contouring as compared with a gold standard in three-dimensional radiotherapy for prostate. Int J Radiat Oncol Biol Phys 2010 Feb 01;76(2):379-385. [doi: 10.1016/j.ijrobp.2009.02.008] [Medline: 19467804]
50. Tan S, Hu A, Wilson T, Ladak H, Haase P, Fung K. Role of a computer-generated three-dimensional laryngeal model in anatomy teaching for advanced learners. J Laryngol Otol 2012 Apr;126(4):395-401. [doi: 10.1017/S0022215111002830] [Medline: 22075619]
51. West N, Konge L, Cayé-Thomasen P, Sørensen MS, Andersen SA. Peak and ceiling effects in final-product analysis of mastoidectomy performance. J Laryngol Otol 2015 Nov;129(11):1091-1096. [doi: 10.1017/S0022215115002364] [Medline: 26391052]
52. Yeung JC, Fung K, Wilson TD. Prospective evaluation of a web-based three-dimensional cranial nerve simulation. J Otolaryngol Head Neck Surg 2012 Dec;41(6):426-436. [doi: 10.2310/7070.2012.00049] [Medline: 23700589]
53. Zaveri PP, Davis AB, O'Connell KJ, Willner E, Aronson Schinasi DA, Ottolini M. Virtual reality for pediatric sedation: a randomized controlled trial using simulation. Cureus 2016 Feb 9;8(2):e486 [FREE Full text] [doi: 10.7759/cureus.486] [Medline: 27014520]
54. Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Cochrane Handbook for Systematic Reviews of Interventions. London, United Kingdom: Wiley; 2011. Presenting results and 'summary of findings' tables URL: https://handbook-5-1.cochrane.org/chapter_11/11_presenting_results_and_summary_of_findings_tables.htm [accessed 2019-01-15] [WebCite Cache ID 75RPkZiwX]

Abbreviations

2D: two-dimensional
3D: three-dimensional
GRADE: Grading of Recommendations, Assessment, Development and Evaluations
MCQ: multiple choice question
OSCE: objective structured clinical examination
SMD: standardized mean difference
VHP: virtual health professional
VP: virtual patient
VR: virtual reality
WHO: World Health Organization

Edited by A Marusic; submitted 29.11.18; peer-reviewed by V Katavic, M Oo; accepted 20.12.18; published 22.01.19.

Please cite as:
Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, Divakar U, Masiello I, Kononowicz AA, Zary N, Tudor Car L
Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration
J Med Internet Res 2019;21(1):e12959
URL: http://www.jmir.org/2019/1/e12959/
doi:10.2196/12959
PMID:30668519

Original Paper

Server-Focused Security Assessment of Mobile Health Apps for Popular Mobile Platforms

Jannis Müthing[1], BSc; Raphael Brüngel[1], BSc; Christoph M Friedrich[1,2], PhD

[1]University of Applied Sciences and Arts Dortmund, Department of Computer Science, Dortmund, Germany

[2]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, Essen, Germany

Corresponding Author:
Christoph M Friedrich, PhD
University of Applied Sciences and Arts Dortmund
Department of Computer Science
Emil-Figge-Straße 42
Dortmund, 44227
Germany
Phone: 49 231 9112 ext 6796
Email: christoph.friedrich@fh-dortmund.de

Abstract

Background: The importance of mobile health (mHealth) apps is growing. Independent of the technologies used, mHealth apps bring more functionality into the hands of users. In the health context, mHealth apps play an important role in providing information and services to patients, offering health care professionals ways to monitor vital parameters or consult patients remotely. The importance of confidentiality in health care and the opaqueness of transport security in apps make the latter an important research subject.

Objective: This study aimed to (1) identify relevant security concerns on the server side of mHealth apps, (2) test a subset of mHealth apps regarding their vulnerability to those concerns, and (3) compare the servers used by mHealth apps with servers used in all domains.

Methods: Server security characteristics relevant to the security of mHealth apps were assessed, presented, and discussed. To evaluate servers, appropriate tools were selected. Apps from the Android and iOS app stores were selected and tested, and the results for functional and other backend servers were evaluated.

Results: The 60 apps tested communicate with 823 servers. Of these, 291 were categorized as functional backend servers, and 44 (44/291, 15.1%) of these received a rating below the A range (A+, A, and A−) by Qualys SSL Labs. A chi-square test was conducted against the number of servers receiving such ratings from SSL Pulse by Qualys SSL Labs. It was found that the tested servers from mHealth apps received significantly fewer ratings below the A range (P<.001). The internationally available apps from the test set performed significantly better than those only available in the German stores (alpha=.05; P=.03). Of the 60 apps, 28 (28/60, 47%) were found using at least one functional backend server that received a rating below the A range from Qualys SSL Labs, endangering confidentiality, authenticity, and integrity of the data displayed. The number of apps that used at least one entirely unsecured connection was 20 (20/60, 33%) when communicating with functional backend servers. It was also found that a majority of apps used advertising, tracking, or external content provider servers. When looking at all nonfunctional backend servers, 48 (48/60, 80%) apps used at least one server that received a rating below the A range.

Conclusions: The results show that although servers in the mHealth domain perform significantly better regarding their security, there are still problems with the configuration of some. The most severe problems observed can expose patient communication with health care professionals, be exploited to display false or harmful information, or used to send data to an app facilitating further damage on the device. Following the recommendations for mHealth app developers, the most regularly observed security issues can be avoided or mitigated.

(J Med Internet Res 2019;21(1):e9818) doi:10.2196/jmir.9818

KEYWORDS

mobile health; mobile apps; data security; computer security; confidentiality; health information technology; servers; data protection

Introduction

Mobile Health Apps

The ubiquitous availability of the internet and mobile devices facilitates new and powerful applications. Mobile health (mHealth) is describing health care–related usage of mobile devices [1]. Although some apps offer user-tailored health information, others facilitate easier communication between patients and health care professionals or offer shop systems for medication sale. mHealth apps have an inherently higher need for security. Besides the importance to protect health data of patients that are collected using apps, by ensuring authenticity of the communication partner and confidentiality of data in transit, information coming from apps must also be protected from unauthorized changes to the content displayed (thus data integrity must be maintained). In contrast to browser-based websites or Web apps, the security of connections used by (native) mobile apps is not transparent to the user. Web browsers display both the use of a secure connection and, importantly, display issues with connection security prominently [2].

In earlier research, client-focused transport security concerns regarding mobile apps were studied [3]; this study focuses on the security of server infrastructure of mHealth apps.

Server Security

The functionality of most apps relies on communication with a remote server over the internet. HTTP is the standard when it comes to client-server communication in the context of mobile apps [4] and offers no security features [5]. The communication through public infrastructure can potentially be observed, modified, or redirected. This endangers the integrity of data displayed by an app and confidentiality of the data sent to and received from a server and also could enable a malicious party to impersonate a server. Furthermore, a publicly reachable server must also guarantee availability [6].

The infrastructure between client and server represents an untrusted medium. Any third party in a privileged position between both communication partners can read and modify all data exchanged. A privileged position in this context is any hop on the path between them (bridges, routers, and gateways) [7]. A common attack vector is a technique called address resolution protocol (ARP) spoofing [7,8]. This enables the attacker to receive all requests normally intended for the router on the local network. Attacks where the third party is located between client and server and can read and modify messages are called man-in-the-middle (MitM) attacks. Another recent example enabling MitM attacks against clients of Wi-Fi Protected Access (WPA) 2 wireless networks is the Key Reinstallation Attack (KRACK) [9]. Older or unpatched protocols such as WPA and wired equivalent privacy (WEP) are also vulnerable to attacks, enabling traffic decryption or MitM setups [10,11]. For attacks on Wi-Fi infrastructure, physical proximity to the target is required. An attacker at least needs to be in the range of the victim's access point. For the KRACK attack, the attacker needs to be close to the victim's device.

To ensure confidentiality and integrity of data sent through an untrusted medium, the Transport Layer Security (TLS) protocol is used [12]. It is an important part of today's internet security infrastructure. TLS was designed for authenticity, integrity, and confidentiality protection of the underlying communication channel by offering secure authentication, data integrity protection, and confidentiality through asymmetric and symmetric cryptography. TLS is situated in the application layer of the transmission control protocol (TCP)/ internet protocol (IP) stack and can wrap and secure HTTP connections. TLS-secured HTTP connections are called HTTP Secure (HTTPS) connections [13].

Given certain server and client configurations, TLS can be set up to offer forward secrecy [14]. This means if secret keys are compromised in the future, past communication stays secure and cannot be decrypted with the compromised credentials alone.

Without this secure wrapper authenticity of a server, confidentiality of information exchanged with a server and the integrity of data sent to an app cannot be guaranteed, and the app might display arbitrary text, pictures, or video data. Because an attacker can potentially feed an app arbitrary input, missing integrity can also be exploited to make the app decode an image, video, or other data that could exploit vulnerabilities in decoders. The Stagefright exploit on Android devices, for example, relied on Android operating system (OS) processing modified media files [15].

TLS and its predecessor the Secure Socket Layer (SSL) protocol are not without security flaws. Since its introduction, multiple vulnerabilities in different layers of the protocols or implementations of the protocols were found and exploited to undermine their security [16]. To keep a server (and therefore patients) secure, special scrutiny and vigilance regarding new threats are required from server operators [17]. It is crucial to react quickly to the publication of new vulnerabilities. A malicious third party only has to test the exploitability of all known vulnerabilities to find a way to attack the server-client communication. Some well-known examples are the Padding Oracle On Downgraded Legacy Encryption (POODLE), Heartbleed, and the recent Return Of Bleichenbacher's Oracle Threat (ROBOT) that was made public in December 2017 [18-20]. To address newly found security issues in SSL and its successor TLS, new versions of the protocol are released regularly. Use of a newer version protects from known security flaws of older versions.

TLS relies on digital certificates to authenticate a server to clients [12,21]. These certificates must be issued (and signed) by certificate authorities (CAs) and have multiple characteristics that must be checked for a certificate to be valid for a given domain. Some characteristics are that a certificate must (1) be issued for the domain requested, (2) have a *valid from* date and must be before the current date, (3) have a *valid until* date in the future, and (4) must not be revoked.

A certificate's revocation status can be checked against a CA's certificate revocation list. As CAs are the roots of the chains of trust, they must operate responsibly. There are some CAs that offer certificate services for free (eg, the Linux foundation's *Let's Encrypt*), whereas others charge for certificates issued by them [22]. The management of the server certificates and

safekeeping of corresponding private keys are also crucial to the security a server can offer.

A general overview of security threats to internet-connected systems can be found at the Open Web Application Security Project (OWASP) [23]. The project collects information, tools, and best practices to avoid common security issues. The OWASP Top 10 and the OWASP mobile Top 10 are of great relevance to the research presented here [24,25]. Although not all vulnerabilities are relevant, the lists were a valuable starting point for the design of the tests.

Prior Work

In prior research, the transport security of mHealth apps from a client's perspective was investigated [3]. The research inspected the data exchanged between iOS and Android client apps and a server and evaluated it under security considerations. This also included the use of TLS and the TLS version. Furthermore, it considered the validation of server certificates by the clients. The study found severe problems with 40% of all 53 tested apps.

Existing literature also evaluated metadata of mHealth apps from Google's Play Store and Apple's App Store [26]. The study did not perform tests or technical analysis. A study of popular mobile apps from China found 97% of apps surveyed provided no information security [27]. The authors limited their investigation to the evaluation of available documentation and availability of auditing reports.

Other research focused on 22 mHealth apps and found that 18 of those apps send data unencrypted over the internet [28].

As mobile apps are not limited to stock HTTP(S) implementations, the study of HTTPS implementations in Android apps is relevant [29]. The publication discusses common flaws in TLS deployments and server configurations.

In the realm of the Internet of Things, existing research analyzed internet-connected toys for children in regard to security and privacy concerns [30]. The authors also bring attention to severe transport security issues of toys when communicating with their backend servers.

Servers have been in use for many years to serve websites and interfaces for internet-based services. The servers' purpose to supply mHealth apps with data and functionality is only one of their most recent use cases. Due to their inherent exposure to public infrastructure (the internet), they always offered an attractive attack surface. As a consequence, knowledge about problems and secure configuration of servers is widely available [5,6,31,32]. An overview of the general landscape of SSL/TLS security on the internet can be observed on Qualys SSL Lab's SSL Pulse website [33].

mHealth apps are getting attention by the media and are covered in regard to treatment of patient data [34]. There is an initiative to build a central place to rate apps in regard to privacy matters (among other criteria) [35]. PrivacyScore has similar objectives and offers a configurable interface to test for a number of security and privacy issues of websites [36].

Existing research mostly considers nontechnical characteristics of mHealth apps, client-side implementations of apps, or solely the use of any encryption at all. In contrast, this research will focus on the configuration of servers used by mHealth apps.

The Methods section will describe how the tested apps were selected. Furthermore, it presents and explains transport security issues for servers and lists the tools used to test for these issues. In the Results section, the tested apps are presented. The test methodology is explained before the aggregated results are listed. These results will be discussed, and common issues will be pointed out.

Methods

App Selection

In prior research, free apps from 3 different European app stores were selected. As differences in behavior between apps from different European countries were not found without loss of generality for this study, only free apps from the German app stores' top lists were chosen [3]. Many apps from the German top-downloaded lists are popular across other nations' app stores. The difference between internationally available and popular apps and apps only in the top lists of the German stores will be discussed in the Results section. To mitigate any platform-dependent bias, apps for Android and iOS are tested.

Relevant Server Security Considerations

HTTP by itself transmits information as clear text without any encryption. It is an application layer protocol and can be secured by being used on top of a secure TLS connection [4,12,13]. TLS and its predecessor SSL are designed to ensure the authenticity of communication partners, confidentiality between parties, and integrity of transmitted data. To achieve this, TLS uses asymmetric cryptography and public key infrastructure (PKI) for authentication and exchange of key material. Symmetric encryption is used for payload data encryption [37,38].

SSL and TLS use version numbers. As earlier versions of the protocol had serious security issues, this paper will take the version of SSL or TLS into consideration [16,18]. While SSL 2.0 is considered insecure because of structural vulnerabilities [39], the POODLE exploit enables third parties to recover plaintext from SSL 3.0-protected traffic [40,41]. Apart from SSL 2.0 and SSL 3.0, newer TLS versions do not have known security vulnerabilities if the server (and client) is properly configured. Lacking a proper configuration, older versions such as TLS 1.0 are vulnerable to an improved POODLE attack and other vulnerabilities [18]. Another TLS 1.0 vulnerability can only be efficiently mitigated by the clients: the Browser Exploit Against SSL/TLS (BEAST) [42]. Although most modern browsers do mitigate the issue, the security of the protocol is still not controllable on the server side. TLS 1.1 and later protocols are not vulnerable to such attacks and can be configured on the server side to use secure ciphers [43]. More recent versions include improvements that are considered more secure. The use of SSL/TLS and the lowest supported version number will be part of the evaluation. We also evaluated support for the recently approved TLS 1.3 (August 2018) and mentioned it in the Results section [44].

To ensure the use of HTTPS and prevent protocol downgrade attacks, HTTP Strict Transport Security (HSTS) can be used [45]. A downgrade attack is designed to get a client to connect to the server using an unsecured HTTP connection. This enables a malicious third party to perform a MitM attack and consequently read and modify sensitive information. For HSTS to be used, the server sends a special header in response to a request. This tells the client to only connect through secure HTTPS connections. For HSTS to work, both client and server have to support it. Although HSTS is most important for browser-based Web apps, many apps include Web components that use platform browser components. In addition, Web-based versions of apps often exist. The presence of HSTS headers in server responses will be listed in the Results section.

To provide a better understanding of the following considerations, a basic understanding of the TLS handshake is required [12]. This description will focus on the most common case of server authentication only.

The client initiates the connection by sending a *client hello* message. This includes its highest supported TLS version, a random value, suggested compression methods, and its supported cipher suites. The server, in turn, answers with the chosen protocol version, cipher suite, compression method, and a random value. In a *certificate message*, the server also sends its certificate. The client now creates a premaster secret, encrypts it with the server's public key, and sends it to the server. Both parties generate the master secret and session keys based on the premaster secret. The client sends a *change cipher spec* message to the server to inform the server that it will use the session key for hashing and message encryption. This is followed by a *client finished* message. The server receives this message, switches to symmetric encryption for further messages, and sends a *server finished* message.

During the TLS handshake, the client uses the servers' public key from its certificate to encrypt a premaster secret. The encryption algorithm is dependent on the negotiated cipher suite. The certificate sent by the server fits the negotiated cipher suite: if, for example, a cipher suite is chosen that includes the *Elliptic Curve Diffie Hellman Ephemeral* (ECDHE) algorithm for key exchange, the certificate must include an elliptic curve (EC) public key [46,47]. As explained in the TLS handshake, the security of the key exchange is essential for the security of the connection. If an adversary is able to decrypt the *premaster* key by brute force, the security of the TLS connection would be compromised. To make this harder, the algorithm used as well as the key length of the public key are important. A commonly used algorithm is the Rivest-Shamir-Adleman (RSA) algorithm [48]. Key sizes vary from 1024 to 4096 bit. It has been shown that a 1024 bit key does not offer sufficient security [49]. Moreover, 2048 bit is the commonly recommended lower limit for RSA keys. The added complexity and negative performance impact during the TLS handshake are disadvantages of the use of longer keys. Newer algorithms such as the EC algorithm do require smaller keys, less computational requirements for clients, and servers while offering equivalent security [47,49]. The key algorithm and length are part of this evaluation.

Another aspect related to the handshake is the selection of the cipher suite. A server has a number of supported cipher suites [12,50]. When the client sends its list of possible cipher suites, the server selects one it supports. The most secure cipher suite should be negotiated between client and server. A server can be configured to have an order of preference for cipher suites [17]. If present, the server will choose the suite highest in priority, which is supported by the client. Whether a server has a preferred order is part of the results of the study because of the importance of the chosen cipher suite for the encryption of user traffic.

For the same reason, the list of supported cipher suites will be considered. Although uncommon, in the worst case, cipher suites may define no encryption at all for the TLS traffic. Other cipher suites define algorithms that can be cryptographically attacked and should not be used anymore. This section does not list all ciphers and their vulnerabilities, but the use of insecure cipher suites is part of the evaluation.

The authentication of the server is based on the server's certificate [12,21,51]. A certificate must fit certain criteria to be considered valid by the client. It must be issued for the requested domain, must not be expired or revoked, and must be trusted. In a PKI, a client trusts root certificates issued by CAs. These CAs use the corresponding private key to sign certificates of servers (or other sub-CAs). When the client verifies the validity of a server certificate, it follows this chain of trust from the server certificate until one of the certificates in the client's trusted root certificates is referenced. Certificate (chain) issues are also part of the research performed for this study.

Older SSL and TLS versions are vulnerable to certain exploits undermining their security. In addition to vulnerabilities of older versions, implementation-dependent issues such as Heartbleed and others are relevant. Heartbleed is an issue in the popular OpenSSL cryptographic software library. It enables an attacker to read memory contents from the server if the library is not patched. Another recently (December 2017) discovered attack uses an issue in RSA implementations and makes the key exchange observable by attackers (ROBOT) [19]. A server that supports RSA for its key exchange, and that is using a vulnerable implementation, is at risk to be attacked [52]. A relative of the BEAST vulnerability discussed earlier also enables attacks against TLS 1.2. It is named the Compression Ratio Info-leak Made Easy (CRIME) and works in conjunction with the use of cookies by protocols that use data compression (such as HTTPS) [53]. It can be used to observe and use a client's authentication cookie to enable further attacks. The vulnerability can be counteracted on the client as well as on the server side. The Browser Reconnaissance and Exfiltration via Adaptive Compression of Hypertext (BREACH) attack is a variant of CRIME and can be exploited to a similar effect [53]. A comprehensive overview of known attacks against TLS can be found as a Request For Comments (RFC) by the Internet Engineering Task Force (IETF) [53]. Vulnerability to known attacks is considered during the tests.

Because the physical location of a server has consequences for applicable law, the server location is considered in this study.

Specifically, this study lists whether servers are inside the European Union (EU), as it is considerably harder for service providers dealing with data belonging to European citizens to host these data outside the EU and still comply with the General Data Protection Regulation (GDPR) [54,55]. Explicit exceptions are GDPR-compliant servers outside the EU covered under the EU-US privacy shield [54,56]. As mentioned in the Limitations section, this study does not differentiate GDPR-compliant servers outside the EU, and servers are simply listed to be located outside EU territory.

Selection of Appropriate Test Tools

The first step of the tests is to find out which of the apps communicate with which backend. Similar to prior research, the BProxy tool was used to facilitate the first test phase [3,57]. This software acts as a proxy between a device running an app and the internet. It is open source and can be found on GitHub [57]. All traffic from the device is analyzed by the proxy, and a report is displayed. As this study is concerned with server-side transport security aspects, only the domain names of servers an app communicates with and the use of unsecured connections are of interest. As the proxy records all traffic to and from the device, filtering is necessary to find out which domains are part of the app's backend. Traffic was inspected to distinguish between app-related servers and other servers. To facilitate correct categorizations, the disconnect service was used and the results were inspected manually, similar to the methodology of a prior publication [3,58].

To test the servers behind the domains, the testssl script and the Qualys SSL Labs test suite were chosen [59,60]. The testssl script uses OpenSSL to perform tests against a targeted server from the local machine. It generates a comma-separated value (CSV) file containing all findings for the domain. The script checks for some relevant characteristics, including all described relevant concerns. The script is actively developed and maintained to contain tests for recently discovered vulnerabilities. At the time of testing, the ROBOT attack was relatively new and already included in the development version of the testssl script [19,59]. The SSL Labs test suite is a Web-based tool for SSL/TLS-related tests. Qualys SSL Labs also offers a command line reference implementation for test automation [61]. The results include similar attributes as the testssl script but importantly assign a graded score A to F. This score is the result of an automated evaluation of the characteristics and vulnerabilities observed. A guide on how this score is calculated and, consequently, how SSL Labs rates the severity of security characteristics can be found on GitHub [62]. This rating guide is updated by Qualys on a regular basis, and a changelog can be found as part of the guide mentioned previously. The rating consists of multiple considerations and includes all but the first item in Table 1: (1) validity and trust of the certificate used by the server, (2) supported protocols (SSL 1.0 up to TLS 1.3), (3) key exchange algorithms supported (older algorithms score lower because of security issues), and (4) cipher suites supported (if no secure, up-to-date cipher suites are supported, the grade will be lower).

Each category scores between 0 and 100. The scores are combined, which results in a single overall score for the server. A 0 in any category will result in an overall score of 0. Although a low score results in an overall lower result, a 0 in a category is indicating a fatal security issue. An example for the score calculation in the supported protocols category according to the Qualys SSL Labs' server rating guide looks like this:

1. Start with the score of the best protocol.
2. Add the score of the worst protocol.
3. Divide the total by 2 [62].

Table 2 lists the scores that Qualys SSL Labs assigns each supported protocol version at the time of writing. A 0 in this category, for example, can only occur if only SSL 2.0 was supported. As this is a long outdated and insecure version, a score of 0 is justified. The other categories are evaluated in a similar manner.

Server configurations that cannot be captured by a score are accounted for by special rules to correct the grade calculated [62]. An example of a server rated B and the scores in each category can be observed in Figure 1.

Table 1. Main considerations evaluated in this study.

Security considerations[a]	Description
Use of secured connections (SSL[b]/TLS[c])	The use of any unsecured connections
SSL/TLS version	Evaluating the supported versions of SSL/TLS
Key exchange support	The cryptographic algorithm used to exchange the keys during the handshake for the following symmetric encryption
Cipher support	The cipher negotiated between client and server dictates what symmetric encryption is applied after the handshake and key exchange
Certificates	The security characteristics TLS offers rely on the server's certificate. Any trust issues here are critical
Vulnerabilities	Certain attacks are based on specific implementations or the absence of a patch on the server
HSTS[d]	Support HSTS can prevent downgrades to HTTP

[a]All but the first one (use of unsecure connections) are tested for by the tools presented in later sections.

[b]SSL: Secure Socket Layer.

[c]TSL: Transport Layer Security.

[d]HSTS: Hypertext Transfer Protocol Strict Transport Security.

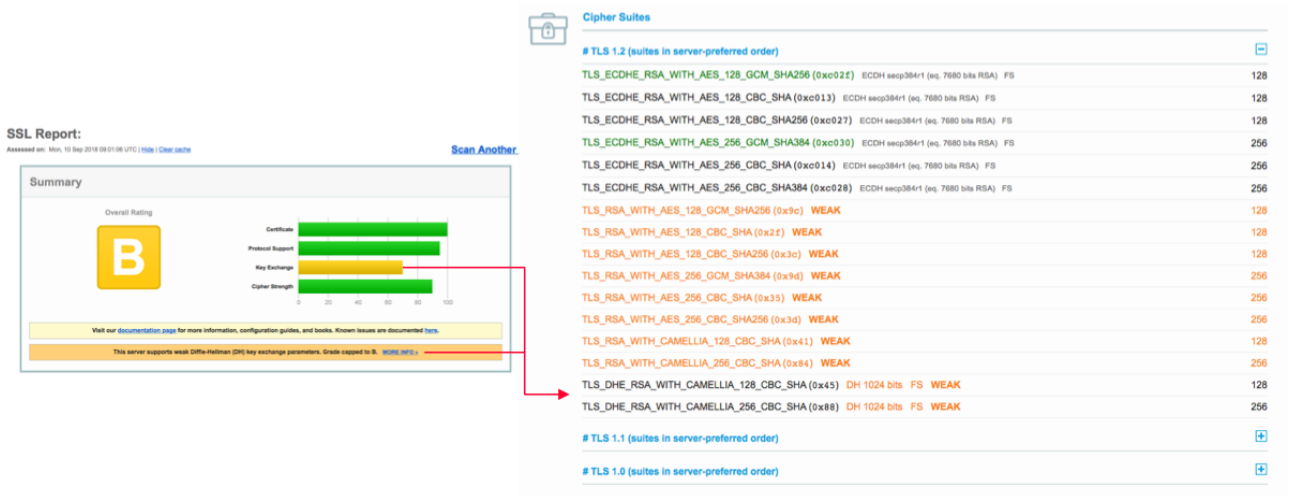
Table 2. Qualys SSL Labs scoring for protocol support.

Protocol	Score (%)
SSL[a] 2.0	0
SSL 3.0	80
TLS[b] 1.0	90
TLS 1.1	95
TLS 1.2	100

[a]SSL: Secure Socket Layer.

[b]TLS: Transport Layer Security.

Figure 1. Exemplary rating and scores of a domain in the test pool. The server was downgraded mainly for offering weak Diffie-Hellmann key exchange. The scores in the distinct categories can be observed to the left. On the right, the offered cipher suites, which include the key exchange algorithms, are listed and marked as weak points.



The overall score, based on a constantly updated set of rules, and the added safety of 2 sources for server security assessment are the reasons why both testssl (Version: 2.9.5dev) and the Qualys SSL Labs suite (Version: 1.32.6) are used. To determine the physical location of a server, a Web-based service was used [63,64]. To aggregate the results, a consistent logic was chosen. For negative observations (such as a rating below the A range containing ratings A+, A, and A− from Qualys SSL Labs), the worst observation in a given category was recorded. For positive observations, the inverse logic was applied: a category of an app was counted as supporting the positive characteristic (such as HSTS and TLS 1.3) when at least one server supported the feature.

Limitations

The tests performed as part of this research focus on transport security and weaknesses in TLS configurations on the server side. The research did not include a full penetration test for every server that was found. Although availability is an important characteristic for a server, this attribute was not part of our tests as testing for denial-of-service (DOS) hardening would require staging test DOS attacks [6]. This would not be legally possible without cooperation of every app developer [65].

As this study focuses on the server side, correct certificate validation in the apps is not tested, neither is the traffic analyzed for leaked information of any kind.

Apps that did not function normally made a test of functions impossible and were removed during app selection. This includes apps that prevented traffic inspection by our test tool. It is possible that these apps used certificate pinning to ensure the servers identity.

The app tests were performed manually. Complete functional coverage cannot be guaranteed. Especially, functionality behind paywalls (in-app purchases or similar facilities) remained untested.

Categorization of servers as functional backends or others was aided by a Web-based service that keeps a list of known advertising and tracking servers but, in part, had to be performed manually [58]. The categorization in uncertain cases was vigilantly checked but might still contain errors.

The location of a server is hard to pin down. It is possible for a server to be inside the EU, although our test using a Web-based service did state a different location. Servers are often used to answer requests from different locales. A request from within the EU might be answered from a server in Ireland, whereas a request from the United States might be handled by a server located in the United States.

A server based outside the EU might fall under the EU-US privacy shield [56]. This allows organizations to store data outside the EU and still comply with applicable EU law [54]. As there is no straightforward way known to the authors to check if a server is part of privacy shield or complies with the GDPR in other ways, this differentiation is not made in this study.

A correct certificate validation depends on a dependable PKI. Research has shown that domain name system (DNS)–based domain validation by CAs is not always dependable and can be abused to have a CA issue certificates for arbitrary domains [66,67]. An attacker in possession of a valid certificate for the domain name requested can impersonate an authentic server even when the client applies correct certificate validation. As this is a CA's responsibility and an app provider has no control in this regard, this issue will not be addressed further in this study.

Results

App Selection

To select a sample set of mHealth apps, the top-downloaded lists of free apps from the Android and iOS app stores are considered as a starting point. The *medical* category was selected as it is most likely to contain mHealth apps. In previous research, similar client security problems in apps from 3 different countries were found [3]. Differences in behavior between apps from stores of different countries were not found. In this study, 30 apps from the German Google Play store and 30 apps from the iOS AppStore are considered (60 apps in total). The top-downloaded lists were generated from the app store analytics site AppAnnie on August 31, 2018 [68]. Top lists for a specific day are available after registration on the website. The *medical* category includes apps that fulfill a broad spectrum of functions. For this reason, apps from both app stores were categorized. All subcategories found in the *medical* categories are listed and defined in Multimedia Appendix 1. The 5 subcategories selected in this study are (1) fertility, pregnancy, and parenthood; (2) drug information, shopping, and compliance; (3) reference and learning; (4) consultation, communication, and interaction; and (5) health, fitness, and monitoring.

To improve variety, the 6 highest positioned apps from these categories and both lists were selected. If a selected app was untestable with test devices, the next most popular app of the same category was selected. This was the case for 9 apps. Of all 60 apps, 26 (26/60, 43%) were in at least one other top 500 list in France, the United Kingdom, or the United States, covering a portion of internationally relevant mHealth apps. The top positions in the stores, developer information, and subcategorization of the apps are visible in Multimedia Appendices 2 and 3.

Performing the Tests

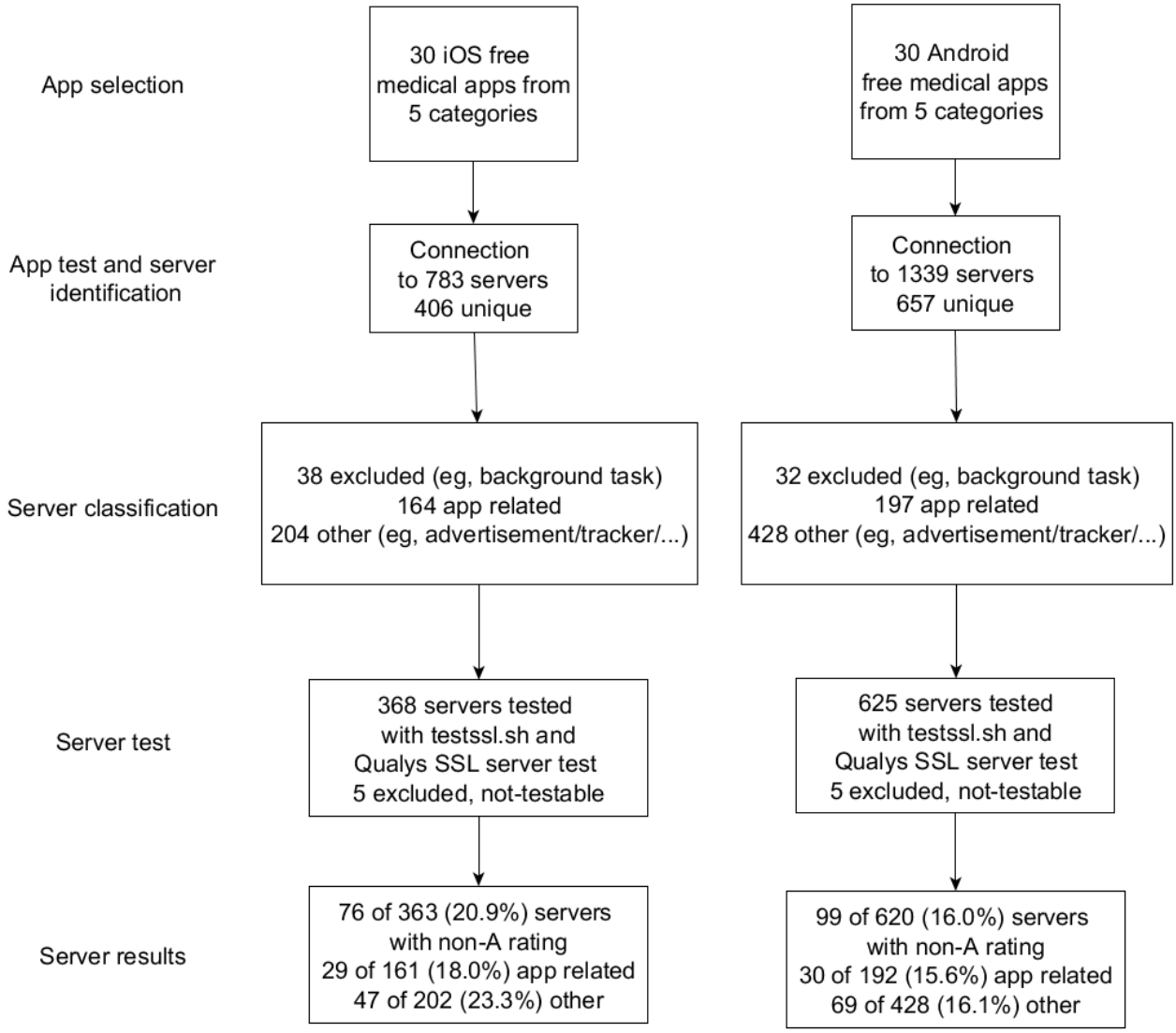
To help parallelize the app testing, iOS and Android apps were tested separately. All apps for both platforms were downloaded from their respective app stores and installed on the devices (iOS 11.4.1 on an iPhone 7 and Android 6.0.1 on a Nexus 7). Before any test, the apps were stopped and restarted. As described previously, the devices are set up to use an HTTP proxy for all HTTP/S traffic. The BProxy tool was used to compile a list of relevant domains.

Many apps communicate with a plethora of endpoints. In addition, the Android or iOS OSs also communicate through HTTP/S connections for multiple purposes, including mail fetching, checking for updates, and sending analytics data. To filter the domains observed during the test of an app, multiple app runs are used to try to distinguish between app traffic and *background* traffic not related to the app. In addition, manual filtering was performed. Service calls from the OS (such as mail server communication) were disregarded. When necessary, an account for the app was created and activated for the tests.

A bash script was used to sequentially test each remaining domain using the testssl script [59]. The Qualys SSL Labs Server Test Web-based tool offers a bulk application programming interface (API) to test multiple domains. Both tools return JavaScript object notation formatted files [59,69]. To reach the objective to generate informative and easy-to-understand results, it was decided to distinguish between backend, advertisement or tracker or analytics, external content, and other servers. The functional backend category includes servers that directly work to supply the app with functional content, perform operations with input from the app, and seem to be under the control of the app developer. All other servers were sorted into the second category of *other* servers, including all servers that serve advertisements, track user activity, and/or are part of an analytics service supplying the app provider with information about app usage. The categorization was performed by evaluating the domain name and analyzing HTTP/S traffic using the BProxy. Categorization as an *other* server was improved using the disconnect service where 311 of the 532 (58.5%) *other* servers were listed [58].

Figure 2 shows a graphical representation of the workflow described. To help evaluate and compile result summaries, the R programming language (version 3.4.2) was used [70]. Negative and positive observations were collected separately. During the compilation of the summaries, a negative or positive observation for a server during the tests was counted as the results for the entire *functional* or *other* backend category of the app. For some analyses, these sets have been combined to obtain results for medical apps in general.

Figure 2. Workflow for tests of mobile health apps. In the app selection phase, the 6 most popular apps from each of the 5 subcategories were selected. In the app test and server identification phase, the traffic between apps and servers was observed and unique servers recorded. The servers were categorized or disregarded as facilitating irrelevant background tasks (server classification phase). The relevant servers were used as the input for the testssl script and the Qualys SSL Labs suite (server test phase). Finally, the results tables were compiled (server results).



Summarized Results

The 60 apps tested communicated with 823 different servers. The distribution of the number of servers apps communicated with can be observed in Table 3. All apps communicated with servers beyond their functional backend. The median number of other servers the apps across both platforms communicated with is 18.5. The median number across Android apps is more than twice (24.5) the amount in comparison with iOS apps (11). In the most remarkable case, an Android app communicated with 82 *other* servers beyond its functional backend.

For easier evaluation of the overall results, the findings were divided into positive and negative results and summarized in Tables 4 and 5. These list the results for both functional as well as other backends. As the functional backends serve requests directly related to an app's functionality, these are the most interesting results.

Detailed results of the security tests can be found in Multimedia Appendices 2 and 3. They list the findings on a per app and per function basis. For further details, please refer to Multimedia Appendices 4 and 5. The tables in these appendices contain the number of connections per apps to servers exhibiting the security characteristics that are part of this study.

Table 3. Minimum, maximum, and median numbers of functional and other backends for iOS and Android apps.

Statistics	Android (functional)	iOS (functional)	Android (others)	iOS (others)	Overall (functional)	Overall (others)
Minimum number of servers	0	1	2	1	0	1
Maximum number of servers	33	21	82	39	33	82
Median number of servers	5	4	24.5	11	4.5	18.5

Table 4. A summarized table of negative results regarding backends of Android and iOS apps. Negative observations are counted for the functional or other category on a per-app basis when it was present in at least one of the apps' servers.

Security issues	Android (functional), n=30	iOS (functional), n=30	Android (others), n=30	iOS (others), n=30	Total (functional), n=60, n (%)	Total (others), n=60, n (%)
Qualys SSL Labs non-A rating	14	14	24	24	28 (47)	48 (80)
Server only offers TLS[a] version <1.2	5	3	0	1	8 (13)	1 (2)
Server without set cipher order	7	5	4	1	12 (20)	5 (8)
Certificate (chain) validation issues present	9	5	14	6	14 (23)	20 (33)
Downgrading vulnerabilities	5	4	8	7	9 (15)	15 (25)
Servers outside the EU[b]	24	21	30	30	45 (75)	60 (100)
Missing forward secrecy support	2	2	1	1	4 (7)	2 (3)
Unsecure connection/s observed	10	10	10	8	20 (33)	18 (30)

[a]TLS: Transport Layer Security.

[b]EU: European Union.

Table 5. A summarized table of positive results regarding backends of Android and iOS apps. One observation of a positive characteristic makes the functional or other category count for the app.

Positive findings	Android (functional), n=30	iOS (functional), n=30	Android (others), n=30	iOS (others), n=30	Total (functional), n=60, n (%)	Total (others), n=60, n (%)
TLS[a] 1.3 support observed	4	5	21	17	9 (15)	38 (63)
HSTS[b] support observed	12	15	28	25	27 (45)	53 (88)

[a]TLS: Transport Layer Security.

[b]HSTS: Hypertext Transfer Protocol Strict Transport Security.

Of the apps tested, 28 (28/60, 47%) used servers where at least one functional backend received a non-A rating from Qualys SSL Labs. In contrast, 48 (48/60, 80%) apps used advertisement, analytics, or external content providers (others) that received a rating below the A range.

Regarding the support of TLS 1.2, only 8 (8/60, 13%) apps used functional backend servers that did not offer TLS 1.2 (3 iOS apps and 5 Android apps). All but one app used only other servers that offered TLS 1.2.

Functional backend servers without a set cipher order were observed when using 12 (12/60, 20%) apps. Other servers without a cipher order were used by 5 (5/60, 8%) apps.

It was found that 14 apps (14/60, 23%) used functional backends that did not offer a valid certificate for the domain requested. This also includes certificates that fail proper validation for any reason (certificate chain issues and domain name mismatch). This issue was also found with other servers for 20 (20/60, 33%) apps.

It was discovered that 9 (9/60, 15%) apps worked with functional backend servers that were downgraded by Qualys SSL Labs because of their vulnerability to an exploit. More apps communicated with other backends that were downgraded (15/60, 25%) apps. Vulnerabilities that lead to a downgrade are related to the support of certain cipher suite or unpatched implementations on the server side [62]. Vulnerability to the POODLE attack was the most observed issue that led to a downgrade. The tables in the Multimedia Appendices 2 and 3 can be consulted for further information.

During the tests, it was found that 45 (45/60, 75%) apps appeared to use some functional backend servers outside the EU. The same is true for all 60 (60/60, 100%) apps regarding other backends.

Forward secrecy was supported by all but 4 (4/60, 7%) apps in at least one functional backend server. The support was observed in at least one other backend server for all but 2 (2/60, 3%) apps.

During the tests, we also recorded entirely unsecured connections by the apps to their servers. This was observed in 20 (20/60, 33%) apps during their communication with a functional backend server and in 18 (18/60, 30%) apps with *other* backend server.

During the evaluation of the test data, some apps showed especially severe issues and are responsible for many of the concerns listed in Table 4. Moreover, 1 app provider, for example, offered 4 apps in total (2 iOS and 2 Android apps). All these apps communicated with backends that did not support an up-to-date TLS version. In addition, it was found that the apps used a mix of unprotected HTTP and protected HTTPS connections to communicate with their backends, further undermining the security of the communication.

It was found that of the 291 functional backend servers, 15.1% received ratings below the A range by Qualys SSL Labs. Of the 532 nonbackend servers tested, 18.8% were rated below the A range. Qualys SSL Labs' SSL Pulse website lists popular security characteristics of servers. They summarize these and also show the percentage of servers receiving a rating below the A range. In the statistics from September 2018, of the 139,849 servers, 37.75% received non-A ratings. A conducted chi-square test (alpha=.05) shows the tested servers of mHealth apps to be significantly better rated than servers observed by SSL Pulse in general ($P<.001$ for both functional backends as well as others) [33].

Chi-square tests were also conducted between each of the subcategories regarding the number of apps that communicated with servers receiving a non-A rating. For these statistical tests, iOS and Android apps as well as functional and other servers were not differentiated. It was found that *reference and learning* apps received a significantly worse rating when tested against both *fertility, pregnancy, and parenthood* (alpha=.05; $P=.02$) and *drug information, shopping, and compliance* (alpha=.05; $P=.01$) apps. The other categories showed no significant differences between each other. Looking at differences between internationally available apps and apps only present in the German top lists, the international apps were found to perform significantly better (alpha=.05; $P=.03$). For this test, any app that was also listed in a top 500 list in the United States, the United Kingdom, or France was considered international.

Looking at regularly contacted domains that can be observed in Table 6, the services behind these top 10 most contacted servers receive data from many of the apps tested and will be able to reconstruct a comprehensive user profile.

Many advertising and analytics companies operate multiple second-level domains. Using the disconnect.me lists revealed, a high number of domains requested belonging to Google's and Facebook's services [58]. Google is very much present in the overview. Almost all (55/60, 92%) apps communicated with servers under a Google API domain, and 8 of the 10 most frequently observed second-level domains are attributable to the company.

Of the 60 apps tested, 17 offered a user-controllable opt-out option for certain advertisement or tracking services. In earlier research, these options were mentioned as desirable [3].

As for positive observations, TLS 1.3 support was observed in 9 (9/60, 15%) apps when communicating with their functional backend servers and 38 (38/60, 63%) with other backend servers. HSTS support was observed in at least one functional backend server for 27 (27/60, 45%) apps and at least once for other backend server in 45 (45/60, 75%) apps. The high percentages in the *others* category is partly a result of the way the results were counted. The number of other servers was regularly higher than the number of functional backend servers, and the positive observations were combined using the Boolean or-conjunction. To gain further insight into how many servers of an app's backend were exhibiting which security characteristic, please refer to Multimedia Appendices 4 and 5.

Table 6. Number of apps that communicated with a subdomain of the second-level domains listed.

Domain	Apps, n=60, n (%)
*.googleapis.com	55 (92)
*.google-analytics.com	46 (77)
*.google.com	38 (63)
*.googleapis.com	37 (62)
*.doubleclick.net	36 (60)
*.gstatic.com	33 (55)
*.crashlytics.com	29 (48)
*.google.de	23 (38)
*.googleadservices.com	23 (38)
*.fbcdn.net	11 (18)

Discussion

Principal Findings

Some backends display problematic characteristics. The most problematic cases exhibit invalid server certificates. In these cases, a client cannot distinguish between a real server certificate and a certificate from a third party attempting a MitM attack. The use of outdated TLS protocol versions can lead to integrity, authenticity, and confidentiality issues with data displayed or sent to the server. The implications vary from app to app. Affected apps did facilitate medical and drug information and interaction checks and patient to health care professional communication. The latter case can expose patient information and enable fabrication of answers from the health care professional to the patient.

In contrast to browsers, apps on mobile platforms have a disadvantage: they can (and do) hide transport security issues from their users. Although private data might be sent to a server without properly securing the connection, a developer can choose to ignore the issues. A browser, on the other hand, displays security warnings when a website can only be connected to using unsecure connections and will make interactions harder for the user [71]. Apple's App Transport Security effort is a step in the right direction but will still allow exceptions for app-specified URLs [72]. It could be the objective of further work to evaluate the possibility for a mobile OS to monitor all traffic from an app and warn the user about insecure connections.

Apps that are localized, available, and successful in more than the German app store suggest bigger organizations with more resources. These organizations can be expected to also devote more resources to the (security) maintenance of their servers. The observation that internationally popular apps performed significantly better validates this expectation.

As discussed in the study by Müthing et al [3], the use of advertisement and tracking services in medical apps can pose as a challenge. These services offer app developers insight into the usage of their app. But the data are also collected by these services for further monetization and data mining [73]. As medical (patient) data are protected under special jurisdiction and should be protected for ethical reasons, the use of any third-party services must be met with scrutiny. For many apps, the tests revealed the use of a great number of servers from various services. The high number of apps that use the same services for advertising or analytics can be problematic. These services can collect user information across multiple apps [73,74]. Another reason to be aware of the use of third-party services is the relatively large number of servers used for advertisements or tracking that received a non-A rating from Qualys SSL Labs (48% of all apps).

Looking at the results in regard to server locations, all (60/60, 100%) apps used *other* servers outside the EU. When looking at the source data, it can be observed that these servers are often related to tracking, analytics, and advertising.

The data also reveal severe issues not directly visible when only considering the server setup: apps were observed using entirely unsecured connections or a mixture of secured and unsecured connections in communication with the same server. Although the backend server might be set up to use state-of-the-art TLS and certificates, this will always undermine potential security and put user data at risk.

The wide support of forward secrecy (functional: 15% and others: 63%) can be seen positively, as it prevents unauthorized decryption of sensitive data in the future and entails a performance burden for attacks on the connection. The high number (functional: 45% and others: 88%) in servers supporting HSTS can help to mitigate against protocol downgrade and cookie hijacking attacks.

During an earlier phase of the preparation of this study—and before the tests discussed so far—40 mHealth apps were tested. The 20 most popular apps from the *medical* category from the German iOS and Android app stores were tested with a very similar methodology as was described in this study. A description of the 40 apps, their categories, and the summarized results can be found in Multimedia Appendix 6. Detailed results per app tested for Android and iOS can be found in Multimedia Appendices 7 and 8, respectively. Some apps were tested in both test runs. Although the earlier results showed a lower number of functional backends receiving a Qualys SSL Labs grade below the A range (28% in early 2018 vs 47% in September 2018), there are also positive developments observable. Vulnerability to the ROBOT attack was observable in a nonfunctional backend in early 2018 but was fixed in the later tests. The relative rise of backends receiving a rating below the A range might be caused by a difference in classification but can also be attributed to the changes in the Qualys SSL Labs rating calculation algorithm [62]. This dynamic is characteristic of the fast-paced nature of the security ecosystem of TLS deployments. Although servers are maintained by some providers, the discovery of new vulnerabilities and the reinvention of old ones leading to new threats make constant vigilance a necessity for server administrators and (mHealth) app providers. The apparent security deterioration of servers by their grades might indicate a slower pace of app providers in keeping their TLS deployments secure.

Common Security Concerns and Recommendations

Common security concerns are listed and summarized below. Prevention, mitigation, or alleviation recommendations are given:

1. Apps were found using a server without valid certificates both for their functional backend as well as for other purposes. This is problematic as this implies missing or erroneous certificate validation in the apps. When a client does not expect valid credentials from the server, any attacker can present equally invalid certificates and impersonate the server. This enables MitM attacks. It is strongly recommended to use a trusted CA and have a valid certificate issued for the domains that are to be secured.
2. A number of apps communicate with servers through unsecured channels. To find and prevent apps and users from using unsecured connections, a server could be configured to use HSTS. In addition, client apps should be

inspected and any unsecured URL schemes should be removed.

3. Insecure server configurations indicated by a low Qualys SSL Labs' rating were found in multiple apps' server backends. This includes a missing set cipher order and the support of vulnerable cipher suites. These vulnerabilities can be exploited to undermine the security supplied by the use of TLS. Insecure server configurations can change over time. Any time a new exploit is discovered and/or is widely exploited, a server operator should update his server's configuration. To keep a server configured as securely as possible, the basic security concerns should be understood [5], server-side security patches installed, and the domains should be tested using a service similar to Qualys SSL Labs' server test [60].
4. Most apps use multiple advertising or analytics servers. Not only does this add to the data and processor time used by apps, for medical apps, but can also be especially problematic as the analytics data can undermine a user's or patient's privacy. A patient looking for pregnancy-related content, for example, might be pregnant. In addition, most of these services appear to be located outside the EU, and most apps used at least one such server that received a rating below the A range by Qualys SSL Labs. Although the use of advertising and analytics services is common in mobile apps, mHealth app developers should thoroughly reconsider the usage of such third-party services and frameworks [73-75]. A possible trade-off could be to offer an opt-out function to the user [3].

Conclusions

Modern mHealth apps from popular subcategories were tested in depth and their behavior was analyzed. Although servers of mHealth apps performed significantly better than servers in general, most apps communicate with a considerable number of different servers by different operators. It was observed that these servers and connections to them are regularly not as secure as connections to the apps' functional backends. The services behind some of these servers (advertisement and app analytics) should also be seen critically in regard to user or patient data protection. Almost half of all apps communicate with functional backends that do not offer a secure TLS setup (non-A Qualys SSL Labs rating).

The most severe problems observed in a small number of apps can expose patient communication with health care professionals, be exploited to display false or harmful information to the user, or used to send data to an app facilitating further damage to the device. These problems include communication through entirely unsecured connections, a mix of secured and unsecured connections, invalid certificates used by servers, certificate chain validation issues, missing support for modern TLS versions, and unpatched vulnerabilities.

As made evident by the comparison of the results discussed in this study with results of previous studies, the security of servers is heavily dependent on the existence and propagation of vulnerabilities. A provider of mHealth apps dealing with (potentially) sensitive information should have an even higher interest in keeping their servers up-to-date and vulnerabilities patched.

The recommendations proposed in the previous section can be used by app developers to improve their transport security setup and prevent putting patients and/or users at risk.

Acknowledgments

The authors would like to thank Obioma Pelka for the help during the proofreading process.

Authors' Contributions

CMF, RB, and JM designed the setup of the tests and chose the tools for testing. CMF and RB performed the tests. RB evaluated the results and compiled the results tables. JM drafted the paper. All authors provided corrections and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Definitions of the subcategories of apps found under the medical category.

[PDF File (Adobe PDF File), 63KB - jmir_v21i1e9818_app1.pdf]

Multimedia Appendix 2

Table containing detailed results and information about the Android apps tested.

[XLSX File (Microsoft Excel File), 14KB - jmir_v21i1e9818_app2.xlsx]

Multimedia Appendix 3

Table containing detailed results and information about the iOS apps tested.

[XLSX File (Microsoft Excel File), 14KB - jmir_v21i1e9818_app3.xlsx]

Multimedia Appendix 4

Table containing the numbers of connections to servers per app and security characteristic for the Android apps tested.

[XLSX File (Microsoft Excel File), 23KB - jmir_v21i1e9818_app4.xlsx]

Multimedia Appendix 5

Table containing the numbers of connections to servers per app and security characteristic for the iOS apps tested.

[XLSX File (Microsoft Excel File), 23KB - jmir_v21i1e9818_app5.xlsx]

Multimedia Appendix 6

App descriptions, apps by categories and summarized results from server security results of popular mHealth apps form early 2018 for both iOS and Android.

[PDF File (Adobe PDF File), 480KB - jmir_v21i1e9818_app6.pdf]

Multimedia Appendix 7

Detailed results from server security results of popular mobile health apps from early 2018 for Android.

[XLSX File (Microsoft Excel File), 8KB - jmir_v21i1e9818_app7.xlsx]

Multimedia Appendix 8

Detailed results from server security results of popular mobile health apps from early 2018 for iOS.

[XLSX File (Microsoft Excel File), 8KB - jmir_v21i1e9818_app8.xlsx]

References

1. Istepanian R, Laxminarayan S, Pattichis CS, editors. M-Health: Emerging Mobile Health Systems. New York: Springer; 2006.
2. Akhawe D, Felt AP. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In: Proceedings of the 22nd USENIX conference on Security. 2013 Aug 14 Presented at: SEC'13; August 14-16, 2013; Washington, DC, United States p. 257-272 URL: https://www.usenix.org/system/files/conference/usenixsecurity13/sec13-paper_akhawe.pdf
3. Müthing J, Jäschke T, Friedrich C. Client-focused security assessment of mHealth apps and recommended practices to prevent or mitigate transport security issues. JMIR Mhealth Uhealth 2017 Oct 18;5(10):e147 [FREE Full text] [doi: 10.2196/mhealth.7791] [Medline: 29046271]
4. Fielding R, Irvine UC, Gettys J, Frystyk H, Masinter L, Leach P, et al. IETF Tools.: IETF; 1999. Hypertext Transfer Protocol--HTTP/1.1 URL: https://tools.ietf.org/html/rfc2616 [accessed 2017-09-18] [WebCite Cache ID 6tZREH58V]
5. Ristic I. Bulletproof SSL and TLS: Understanding and Deploying SSL/TLS and PKI to Secure Servers and Web Applications. London: Feisty Duck; 2018.
6. Perrin C. TechRepublic. 2008. The CIA Triad URL: https://www.techrepublic.com/blog/it-security/the-cia-triad/ [accessed 2018-05-11] [WebCite Cache ID 6zJlBN5yg]
7. Plummer D. IETF Tools. 1982 Nov. Ethernet Address Resolution Protocol or Converting Network Protocol Addresses to 48bit Ethernet Address for Transmission on Ethernet Hardware URL: https://tools.ietf.org/html/rfc826 [accessed 2018-11-30] [WebCite Cache ID 74IRcZoIS]
8. Ramachandran V, Nandi S. Detecting ARP Spoofing: An Active Technique. In: Proceedings of the International Conference on Information Systems Security.: Springer; 2005 Presented at: ICISS'05; December 19-21, 2005; Kolkata, India p. 239-250. [doi: 10.1007/11593980_18]
9. Vanhoef M, Piessens F. Key Reinstallation Attacks: Forcing Nonce Reuse in WPA2. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security.: ACM Press; 2017 Presented at: CCS '17; October 30-November 03, 2017; Dallas, Texas, USA p. 1313-1328. [doi: 10.1145/3133956.3134027]
10. Sepehrdad P, Susil P, Vaudenay S, Vuagnoux M. Smashing WEP in a Passive Attack. Berlin, Germany: Springer; 2013 Presented at: International Workshop on Fast Software Encryption; March 11, 2013; Washington, DC, USA p. 155-178. [doi: 10.1007/978-3-662-43933-3_9]
11. Cassola A, Robertson W, Kirda E, Noubir G. A Practical, Targeted, and Stealthy Attack Against WPA Enterprise Authentication. 2013 Presented at: NDSS Symposium 2013; February 24, 2013; San Diego, CA, USA.

12. Dierks T, Allen C. The TLS Protocol Version 1.0. 1999. URL: https://www.rfc-editor.org/rfc/pdfrfc/rfc2246.txt.pdf [accessed 2018-05-11] [WebCite Cache ID 6zKmXnHw7]
13. Rescorla E. HTTP Over TLS. 2000. URL: https://www.rfc-editor.org/rfc/pdfrfc/rfc2818.txt.pdf [accessed 2018-05-11] [WebCite Cache ID 6zKaCoz7G]
14. Huang LS, Adhikarla S, Boneh D, Jackson C. An experimental study of TLS forward secrecy deployments. IEEE Internet Comput 2014 Nov;18(6):43-51. [doi: 10.1109/MIC.2014.86]
15. Drake JJ. Stagefright: An Android Exploitation Case Study. In: 10th USENIX Workshop on Offensive Technologies (WOOT). 2016 Presented at: WOOT '16; August 8-9, 2016; Austin, TX, USA.
16. Meyer C, Schwenk J. International Association for Cryptologic Research.: IACR Cryptology ePrint Archive; 2013. Lessons Learned From Previous SSL/TLS Attacks A Brief Chronology Of Attacks And Weaknesses URL: https://eprint.iacr.org/2013/049.pdf [accessed 2018-05-11] [WebCite Cache ID 6zL5btUjI]
17. Stanek M. Comenius University.: arXiv; 2017 Aug 24. Secure by default - the case of TLS URL: https://arxiv.org/pdf/1708.07569.pdf [accessed 2018-11-30] [WebCite Cache ID 74JyxErWt]
18. Möller B, Duong T, Kotowicz K. Google. 2014. This POODLE Bites: Exploiting The SSL 3.0 Fallback URL: https://www.openssl.org/~bodo/ssl-poodle.pdf [accessed 2018-05-11] [WebCite Cache ID 6zKcsqBIb]
19. Böck H, Somorovsky J, Young C. Return Of Bleichenbacher's Oracle Threat (ROBOT). In: Proceedings of the 27th USENIX Conference on Security Symposium.: USENIX Association; 2018 Presented at: SEC'18; August 15-17, 2018; Baltimore, MD, USA.
20. Synopsys, Inc. 2014. The Heartbleed Bug URL: http://heartbleed.com [accessed 2018-01-22] [WebCite Cache ID 6wemwGqYP]
21. Housley R, Ford W, Polk W, Solo D. Internet X.509 Public Key Infrastructure Certificate and CRL Profile. 1999. URL: https://www.rfc-editor.org/rfc/pdfrfc/rfc2459.txt.pdf [accessed 2018-05-11] [WebCite Cache ID 6zKcZKZNa]
22. Let's Encrypt. San Francisco, CA, USA: Internet Security Research Group (ISRG); 2018. URL: https://letsencrypt.org/ [accessed 2018-05-11] [WebCite Cache ID 6zKUHrmJn]
23. Open Web Application Security Project (OWASP). 2018. URL: https://www.owasp.org/index.php/Main_Page [accessed 2018-01-22] [WebCite Cache ID 6wen4anph]
24. Open Web Application Security Project. 2017. Top 10-2017 Top 10 URL: https://www.owasp.org/index.php/Top_10_2017-Top_10 [accessed 2018-05-11] [WebCite Cache ID 6zKcCrm8S]
25. Open Web Application Security Project. 2016. Mobile Top 10 2016-Top 10 URL: https://www.owasp.org/index.php/Mobile_Top_10_2016-Top_10 [accessed 2017-02-09] [WebCite Cache ID 6o9JIoR8V]
26. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
27. Hsu J, Liu D, Yu YM, Zhao HT, Chen ZR, Li J, et al. The top Chinese mobile health apps: a systematic investigation. J Med Internet Res 2016 Dec 29;18(8):e222 [FREE Full text] [doi: 10.2196/jmir.5955] [Medline: 27573724]
28. He D, Naveed M, Gunter CA, Nahrstedt K. Security concerns in Android mHealth apps. AMIA Annu Symp Proc 2014;2014:645-654 [FREE Full text] [Medline: 25954370]
29. Wei X, Wolf M. A survey on HTTPS implementation by Android apps: issues and countermeasures. Prog Adv Comput Intell Eng 2017 Jul;13(2):101-117 [FREE Full text] [doi: 10.1016/j.aci.2016.10.001]
30. Chu G, Apthorpe N, Feamster N. IEEE Internet Things J. 2018. Security and privacy analyses of internet of things children's toys URL: https://arxiv.org/pdf/1805.02751.pdf [accessed 2018-12-06] [WebCite Cache ID 74Si2UYZc]
31. Oppliger R. SSL and TLS: Theory and Practice, Second Edition. Norwood, MA: Artech House; 2016.
32. Jiang S, Smith S, Minami K. Securing Web servers against insider attack. In: Proceedings of the 17th Annual Computer Security Applications Conference. 2001 Dec 10 Presented at: ACSAC '01; December 10-14, 2001; New Orleans, Louisiana, USA p. 265-276. [doi: 10.1109/ACSAC.2001.991542]
33. Qualys SSL Labs. 2018. SSL Pulse URL: https://www.ssllabs.com/ssl-pulse/ [accessed 2018-01-22] [WebCite Cache ID 6wepClCQS]
34. Zeit. 2018. This app does not help me fall asleep [in German] URL: https://www.zeit.de/wissen/gesundheit/2018-09/psychische-gesundheit-apps-moodpath-schlafstoerungen-erfahrung [accessed 2018-11-30] [WebCite Cache ID 74Jy0oH3g]
35. HealthOn. HealthOn App Code of Ethics for Health Apps [in German] URL: https://www.healthon.de/ehrenkodex [accessed 2018-10-29] [WebCite Cache ID 73XM1QpoP]
36. Maass M, Wichmann P, Pridöhl H, Herrmann D. PrivacyScore: Improving PrivacySecurity via Crowd-Sourced Benchmarks of Websites. In: Schweighofer E, Leitold H, Mitrakas A, Rannenberg K, editors. Privacy Technologies and Policy. Cham: Springer International Publishing; 2017:178-191.
37. Maurer U. Modelling a Public-Key Infrastructure. In: Proceedings of the 4th European Symposium on Research in Computer Security: Computer Security. UK: Springer-Verlag; 1996 Presented at: ESORICS '96; September 25-27, 1996; Rome, Italy p. 325-350 URL: https://dl.acm.org/citation.cfm?id=699185 [doi: 10.1007/3-540-61770-1_45]
38. Housley R, Ford W, Polk W, Solo D. Internet X.509 Public Key Infrastructure Certificate and CRL Profile.: IETF; 1999 Jan. URL: https://www.rfc-editor.org/rfc/rfc2459.txt [accessed 2018-11-30] [WebCite Cache ID 74JyCVbS7]

39. Turner S, Polk T. Prohibiting Secure Sockets Layer (SSL) Version 2.0. 2011 Mar. URL: https://www.rfc-editor.org/rfc/pdfrfc/rfc6176.txt.pdf [accessed 2018-05-11] [WebCite Cache ID 6zKnAWzH3]
40. Barnes R, Thomson M, Pironti A, Langley A. Deprecating Secure Sockets Layer Version 3.0. 2015. URL: https://www.rfc-editor.org/rfc/pdfrfc/rfc7568.txt.pdf [accessed 2018-05-11] [WebCite Cache ID 6zKeVo47b]
41. Rizzo J, Duong T. Practical Padding Oracle Attacks. In: Proceedings of the 4th USENIX conference on Offensive technologies. 2010 Presented at: WOOT'10; August 9, 2010; Berkeley, CA, USA p. 1-8 URL: http://dl.acm.org/citation.cfm?id=1925004.1925008
42. Ristic I. Qualys SSL Labs. 2013. Is BEAST Still a Threat? URL: https://blog.qualys.com/ssllabs/2013/09/10/is-beast-still-a-threat [accessed 2018-01-22] [WebCite Cache ID 6wepLFf5q]
43. Sheffer Y, Holz R, Saint-Andre P. Summarizing Known Attacks on Transport Layer Security (TLS) and Datagram TLS (DTLS). 2015. URL: https://www.rfc-editor.org/rfc/pdfrfc/rfc7457.txt.pdf [accessed 2018-05-11] [WebCite Cache ID 6zKac2MmJ]
44. Rescorla E. The Transport Layer Security (TLS) Protocol Version 1.3.: IETF; 2018 Aug. URL: https://www.rfc-editor.org/rfc/rfc8446.txt [accessed 2018-11-30] [WebCite Cache ID 74JsSwFHr]
45. Hodges J, Jackson C, Barth A. HTTP Strict Transport Security (HSTS).: IETF; 2012 Nov. URL: https://www.rfc-editor.org/rfc/rfc6797.txt [accessed 2018-11-30] [WebCite Cache ID 74JscC3j5]
46. Barker E, Chen L, Roginsky A, Smid M. Recommendation for Pair-Wise Key Establishment Schemes Using Discrete Logarithm Cryptography. In: NIST special publication 800-56A Revision 2. Gaithersburg, MD, USA: NIST; May 2013.
47. Seroussi G. Elliptic curve cryptography. : IEEE; 1999 Jun 27 Presented at: Information Theory and Networking Workshop; 27 June-1 July 1999; Metsovo, Greece p. 41. [doi: 10.1109/ITNW.1999.814351]
48. Rivest RL, Shamir A, Adleman LM. Cryptographic communications system and method. Washington, DC, USA: U.S. Patent and Trademark Office; 1983.
49. Barker E, Roginsky A. NIST Special Publication 800-131A Revision 1. Gaithersburg, MD, USA: NIST; 2015. Transitions: Recommendation for Transitioning the Use of Cryptographic Algorithms and Key Lengths URL: https://www.nist.gov/publications/transitions-recommendation-transitioning-use-cryptographic-algorithms-and-key-lengths-0 [WebCite Cache ID 75cQFIPp5]
50. Dierks T, Rescorla E. The Transport Layer Security (TLS) Protocol Version 1.2.: IETF; 2008 Aug. URL: https://www.rfc-editor.org/rfc/rfc5246.txt [accessed 2018-11-30] [WebCite Cache ID 74JtLXOrt]
51. Maurer U. Modelling a public-key infrastructure. In: Proceedings of the 4th European Symposium on Research in Computer Security: Computer Security. Berlin, Heidelberg: Springer; 1996 Presented at: ESORICS 1996; September 25-27, 1996; Rome, Italy p. 325-350. [doi: 10.1007/3-540-61770-1_45]
52. Böck H, Somorovsky J, Young C. Robot Attack. 2017. The ROBOT Attack - Return Of Bleichenbacher's Oracle Threat URL: https://robotattack.org [accessed 2018-10-03] [WebCite Cache ID 72toEp1MS]
53. Sheffer Y, Holz R, Saint-Andre P. Summarizing Known Attacks on Transport Layer Security (TLS) and Datagram TLS (DTLS).: IETF; 2015 Feb. URL: https://www.rfc-editor.org/rfc/rfc7457.txt [accessed 2018-11-30] [WebCite Cache ID 74JuA4NmD]
54. Official Journal of the European Union. Brussels, Belgium: European Commission; 2016. REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) URL: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=DE, [accessed 2018-01-22] [WebCite Cache ID 6wepXOTzy]
55. European Commission.: European Commission; 2016. Privacy Code of Conduct on mobile health apps URL: https://ec.europa.eu/digital-single-market/en/privacy-code-conduct-mobile-health-apps [accessed 2018-01-22] [WebCite Cache ID 6wepa89cD]
56. Official Journal of the European Union. Brussels, Belgium: European Commission; 2016 Apr 27. COMMISSION IMPLEMENTING DECISION (EU) 2016/1250 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the adequacy of the protection provided by the EU-U.S. Privacy Shield URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016D1250&from=EN [accessed 2018-11-30] [WebCite Cache ID 74JuwyjwB]
57. Müthing J. Github. BProxy URL: https://github.com/j4nnis/bproxy [accessed 2018-01-22] [WebCite Cache ID 6weprpw1R]
58. Disconnect. 2018. URL: https://disconnect.me [accessed 2018-09-28] [WebCite Cache ID 72lxKBy6i]
59. Wetter D. Test SSL. 2018. Testing TLS/SSL encryption URL: https://testssl.sh [accessed 2018-01-22] [WebCite Cache ID 6wepvH2Jd]
60. Qualys SSL Labs. 2018. SSL Server Test URL: https://www.ssllabs.com/ssltest/ [accessed 2018-01-22] [WebCite Cache ID 6wepy1keE]
61. Qualys SSL Labs. 2018. ssllabs-scan URL: https://github.com/ssllabs/ssllabs-scan [accessed 2018-01-22] [WebCite Cache ID 6weq0ynxY]
62. Qualys SSL Labs. 2017 May 08. Qualys SSL Server Rating Guide URL: https://github.com/ssllabs/research/wiki/SSL-Server-Rating-Guide [accessed 2018-01-22] [WebCite Cache ID 6weq5XAMB]
63. IP-API. 2018. IP Geolocation API URL: http://ip-api.com [accessed 2018-11-30] [WebCite Cache ID 74JwHDsIj]

64. Shavitt Y, Zilberman N. A geolocation databases study. IEEE J Select Areas Commun 2011 Dec;29(10):2044-2056. [doi: 10.1109/JSAC.2011.111214]
65. Rasch M. Security Current.: security current; 2013 Nov 26. Legal Issues in Penetration Testing URL: https://securitycurrent.com/legal-issues-in-penetration-testing/ [accessed 2018-09-29] [WebCite Cache ID 72nS3BAlU]
66. Son S, Shmatikov V. The Hitchhiker's Guide to DNS Cache Poisoning. Berlin, Heidelberg: Springer; 2010 Presented at: International Conference on Security and Privacy in Communication Systems; September 7-9, 2010; Singapore, Singapore p. 466-483. [doi: 10.1007/978-3-642-16161-2_27]
67. Herzberg A, Shulman H. Security of Patched DNS. Heidelberg, Berlin: Springer; 2012 Presented at: European Symposium on Research in Computer Security (ESORICS); September 10-12, 2012; Pisa, Italy p. 271-288. [doi: 10.1007/978-3-642-33167-1_16]
68. App Annie. 2018. URL: https://www.appannie.com [accessed 2018-11-30] [WebCite Cache ID 74JxBE6BG]
69. Qualys SSL Labs. 2018. SSL Labs APIs URL: https://www.ssllabs.com/projects/ssllabs-apis/index.html [accessed 2018-10-29] [WebCite Cache ID 73XMSSddH]
70. R Core Team. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. The R Project for Statistical Computing URL: https://www.r-project.org/ [accessed 2018-12-01] [WebCite Cache ID 74JxPQXSy]
71. Reis C, Barth A, Pizano C. Browser security: lessons from Google Chrome. Commun ACM 2009 Aug 01;52(8):45-49. [doi: 10.1145/1536616.1536634]
72. Apple. Cupertino, CA, USA: Apple; 2018. Cocoa Keys URL: https://developer.apple.com/library/archive/documentation/General/Reference/InfoPlistKeyReference/Articles/CocoaKeys.html#//apple_ref/doc/uid/TP40009251-SW33, [accessed 2018-09-10] [WebCite Cache ID 72nSFR3OB]
73. Razaghpanah A, Nithyanand R, Vallina-Rodriguez N, Sundaresan S, Allman M, Kreibich C, et al. Apps, Trackers, Privacy, and Regulators-A Global Study of the Mobile Tracking Ecosystem. 2018 Presented at: Network and Distributed Systems Security (NDSS) Symposium; February 18-21, 2018; San Diego, CA, USA. [doi: 10.14722/ndss.2018.23353]
74. Binns R, Lyngs U, van Kleek M, Zhao J, Libert T, Shadbolt N. Third Party Tracking in the Mobile Ecosystem. In: Proceedings of the 10th ACM Conference on Web Science. New York, NY, USA: ACM; 2018 Presented at: WebSci '18; May 27-30, 2018; Amsterdam, NL p. 23-31. [doi: 10.1145/3201064.3201089]
75. Thurm S, Kane YI. The Wall Street Journal.: The Wall Street Journal; 2010 Dec 17. Your Apps Are Watching You: A WSJ Investigation finds that iPhone and Android apps are breaching the privacy of smartphone users URL: https://www.wsj.com/articles/SB10001424052748704694004576020083703574602 [accessed 2018-09-29] [WebCite Cache ID 72nLyoV6U]

Abbreviations

API: application programming interface
ARP: address resolution protocol
BEAST: Browser Exploit Against Secure Socket Layer/Transport Layer Security
CA: certificate authority
CRIME: Compression Ratio Info-leak Made Easy
CSV: comma-separated value
DNS: domain name system
DOS: denial-of-service
EC: elliptic curve
ECDHE: Elliptic Curve Diffie Hellman Ephemeral
EU: European Union
GDPR: General Data Protection Regulation
HSTS: Hypertext Transfer Protocol Strict Transport Security
HTTPS: HTTP Secure
IETF: Internet Engineering Task Force
IP: internet protocol
KRACK: Key Reinstallation Attack
mHealth: mobile health
MitM: man-in-the-middle
OS: operating system
OWASP: Open Web Application Security Project
PKI: Public Key Infrastructure
POODLE: Padding Oracle On Downgraded Legacy Encryption
ROBOT: Return Of Bleichenbacher's Oracle Threat
RSA: Rivest-Shamir-Adleman

SSL: Secure Socket Layer
TCP: transmission control protocol
TLS: Transport Layer Security
WEP: wired equivalent privacy
WPA: Wi-Fi Protected Access

Edited by M Stanley, E Perakslis; submitted 22.01.18; peer-reviewed by T Dehling, M Abdelhamid, S Albakri, A Ferreira; comments to author 08.05.18; revised version received 03.10.18; accepted 01.11.18; published 23.01.19.

Please cite as:
Müthing J, Brüngel R, Friedrich CM
Server-Focused Security Assessment of Mobile Health Apps for Popular Mobile Platforms
J Med Internet Res 2019;21(1):e9818
URL: https://www.jmir.org/2019/1/e9818/
doi:10.2196/jmir.9818
PMID:30672738

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org